The Astrophysical Journal, 905:93 (20pp), 2020 December 20 https://doi.org/10.3847/1538-4357/abc42b

Photometric Classification of 2315 Pan-STARRS1 Supernovae with Superphot

Griffin Hosseinzadeh[1], Frederick Dauphin[1,2,3], V. Ashley Villar[1,4,18], Edo Berger[1], David O. Jones[5], Peter Challis[1], Ryan Chornock[6], Maria R. Drout[7,8], Ryan J. Foley[5], Robert P. Kirshner[1,9], Ragnhild Lunnan[10], Raffaella Margutti[6], Dan Milisavljevic[11], Yen-Chen Pan[12], Armin Rest[3,13], Daniel M. Scolnic[14], Eugene Magnier[15], Nigel Metcalfe[16], Richard Wainscoat[15], and Christopher Waters[17]

[1] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138-1516, USA; griffin.hosseinzadeh@cfa.harvard.edu
[2] Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3815, USA
[3] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218-2410, USA
[4] Department of Astronomy, Columbia University, New York, NY 10027-6601, USA
[5] Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064-1077, USA
[6] Center for Interdisciplinary Exploration and Research in Astrophysics and Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3112, USA
[7] David A. Dunlap Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, Ontario, M5S 3H4 Canada
[8] Observatories of the Carnegie Institute for Science, 813 Santa Barbara Street, Pasadena, CA 91101-1232, USA
[9] Gordon and Betty Moore Foundation, 1661 Page Mill Road, Palo Alto, CA 94304-1209, USA
[10] Oskar Klein Centre, Department of Astronomy, Stockholm University, Albanova University Centre, SE-106 91 Stockholm, Sweden
[11] Department of Physics and Astronomy, Purdue University, 525 Northwestern Avenue, West Lafayette, IN 47907-2036, USA
[12] Graduate Institute of Astronomy, National Central University, 300 Jhongda Road, Zhongli, Taoyuan, 32001, Taiwan
[13] Department of Physics and Astronomy, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA
[14] Department of Physics, Duke University, Campus Box 90305, Durham, NC 27708, USA
[15] Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822-1839, USA
[16] Department of Physics, Durham University, South Road, Durham, DH1 3LE, UK
[17] Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08540-7219, USA
Received 2020 August 11; revised 2020 October 2; accepted 2020 October 20; published 2020 December 17

Abstract

The classification of supernovae (SNe) and its impact on our understanding of explosion physics and progenitors have traditionally been based on the presence or absence of certain spectral features. However, current and upcoming wide-field time-domain surveys have increased the transient discovery rate far beyond our capacity to obtain even a single spectrum of each new event. We must therefore rely heavily on photometric classification—connecting SN light curves back to their spectroscopically defined classes. Here, we present Superphot, an open-source Python implementation of the machine-learning classification algorithm of Villar et al., and apply it to 2315 previously unclassified transients from the Pan-STARRS1 Medium Deep Survey for which we obtained spectroscopic host-galaxy redshifts. Our classifier achieves an overall accuracy of 82%, with completenesses and purities of >80% for the best classes (SNe Ia and superluminous SNe). For the worst performing SN class (SNe Ibc), the completeness and purity fall to 37% and 21%, respectively. Our classifier provides 1257 newly classified SNe Ia, 521 SNe II, 298 SNe Ibc, 181 SNe IIn, and 58 SLSNe. These are among the largest uniformly observed samples of SNe available in the literature and will enable a wide range of statistical studies of each class.

Unified Astronomy Thesaurus concepts: Supernovae (1668); Astrostatistics (1882); Light curve classification (1954)

Supporting material: data behind figure, machine-readable tables

1. Introduction

Starting with Minkowski (1941), supernovae (SNe) have been classified on the basis of their spectra, with hydrogen-poor events being labeled Type I and hydrogen-rich events Type II. Uomoto & Kirshner (1985) and Wheeler & Levreault (1985) later separated Type Ia SNe (SNe Ia), which show silicon absorption (as well as a secondary infrared light-curve peak; Elias et al. 1985), from the remaining SNe I, which were later divided into helium-rich SNe Ib and helium-poor SNe Ic (Wheeler & Harkness 1986). Similarly, Schlegel (1990) separated SNe IIn, which show narrow hydrogen emission lines, from the remaining SNe II. More recently, Quimby et al. (2011) and Gal-Yam (2012) defined a class of superluminous SNe (SLSNe) that are 10×–100× more luminous than the aforementioned classes; though originally identified photometrically, hydrogen-poor SLSNe are now considered a spectroscopic class (Quimby et al. 2018; Gal-Yam 2019). Subsequent authors have further subdivided all of the aforementioned classes (see Gal-Yam 2016 for a review).

With the advent of high-étendue time-domain facilities like the Panoramic Survey Telescope and Rapid Response System 1 (Pan-STARRS1; Chambers et al. 2016) and Zwicky Transient Facility (Bellm et al. 2019), as well as the upcoming Vera C. Rubin Observatory (Ivezić et al. 2019), the transient discovery rate has far exceeded the worldwide capacity for spectroscopic classification. We must therefore rely on photometric classification methods, despite the fact that the classes are defined spectroscopically.

Many previous attempts at photometric classification have focused only on separating SNe Ia from all other SN classes, for the purpose of measuring cosmological parameters (Riess et al. 2004a, 2004b; Möller et al. 2016; Kimura et al. 2017). This is a somewhat easier problem because of the relative photometric uniformity of SNe Ia. Other attempts have relied on simulated data for their training sets (Richards et al. 2012; Charnock & Moss 2017; Kimura et al. 2017; Boone 2019;

[18] Simons Junior Fellow.

Ishida et al. 2019; Muthukrishna et al. 2019). The use of simulated light curves implies that we understand the full diversity of explosive transients, although it is a logical way forward in the absence of large, uniformly observed photometric data sets.

Villar et al. (2019, hereafter V19) presented a method for photometric classification that consists of (1) fitting a highly flexible analytical model to the observed light curve, (2) extracting features from that model light curve, and (3) using supervised machine learning to classify the SN based on those features. V19 tested 24 different pipelines—consisting of 4 different methods of feature extraction, 2 different methods for balancing the classes in the training set, and 3 different machine-learning algorithms—that they trained on the spectroscopically classified transients in the Pan-STARRS1 Medium Deep Survey (PS1-MDS; Chambers et al. 2016). This is the largest real data set that has been used to train a multiclass classifier.

Here, we present an open-source Python implementation of this method and apply it to 2315 previously unclassified transients from the same survey for which we obtained spectroscopic host-galaxy redshifts. The code is available on GitHub and Zenodo (Hosseinzadeh & Dauphin 2020) and listed on the Python Package Index under the name Superphot.[19,20]

For the purposes of this work, we restrict our classification to the following five broad labels, adopting only slight refinements to the definitions discussed above that are already in common use:

1. "SN Ia" refers to hydrogen-poor SNe that show silicon absorption and a secondary infrared light-curve peak.
2. "SN Ibc" refers to all normal-luminosity, hydrogen-poor SNe that are not SNe Ia. Due to the small number of these in our training sample, we do not distinguish between SNe Ib and SNe Ic.
3. "SN IIn" refers to SNe that show narrow hydrogen emission lines in their spectra throughout their evolution.
4. "SN II" refers to all hydrogen-rich SNe that are not SNe IIn. We do not distinguish between SNe IIL and SNe IIP, because it is not clear that these populations are separate (Sanders et al. 2015; Valenti et al. 2016).
5. "SLSN" refers only to hydrogen-poor SLSNe. Hydrogen-rich SLSNe are included in SNe IIn.

In Section 2, we briefly present the photometric data set we use for training and classification, as well as the spectroscopic data set used to determine the host-galaxy redshifts. In Section 3, we describe the implementation of the algorithm. In Section 4, we apply our algorithm to the unclassified light curves and assess its performance. In Section 5, we discuss its utility for current and future time-domain surveys. In Section 6, we conclude with some lessons learned from this case study in photometric classification.

2. Data Set

Pan-STARRS1 is a 1.8 m telescope near the summit of Haleakalā Hawai'i, equipped with a 1.4 gigapixel camera with a 7 deg^2 field of view (Chambers et al. 2016). PS1-MDS ran from 2009 July to 2014 July using 25% of the observing time on Pan-STARRS1 and consisted of 10 deep-drilling fields with a three-day cadence in any of five bands: *grizy* (Chambers et al. 2016). Images were processed and calibrated using the Image Processing Pipeline (Magnier et al. 2020a, 2020b, 2020c; Waters et al. 2020). With the exception of *y*, PS1-MDS reached typical depths of 23.3 mag per visit (Chambers et al. 2016); the *y* filter has lower throughput and cadence, so we exclude it from our analysis. As such, our data set is very similar to the proposed depth and cadence of the Rubin Observatory Legacy Survey of Space and Time (LSST), but with ∼0.1% of the time-integrated sky coverage (Ivezić et al. 2019).

Over the course of PS1-MDS, we detected a total of 5243 SN-like transients[21] using `photpipe` (Rest et al. 2005, 2014). These light curves are presented in a companion paper by Villar et al. (2020a, hereafter V20) and are available on Zenodo (Villar et al. 2020b). We obtained spectroscopic classifications of 573 of these transients. The 557 SNe belonging to one of the five classes listed in Section 1 comprise our "training set." The 17 SLSNe in our data set were previously published by Lunnan et al. (2018), and 76 of the SN II light curves were analyzed by Sanders et al. (2015). The remaining 16 spectroscopically classified transients belong to less common classes. Because these are too few in number to be used as training samples, we exclude them from our analysis except to explore how they are labeled by our classifier (Section 5.2).

We also obtained host-galaxy spectra for 4233 transients, 3600 of which were not classified spectroscopically: 3434 from MMT, 324 from the Anglo-Australian Telescope (AAT), 301 from WIYN, 169 from the Sloan Digital Sky Survey (SDSS; Ahumada et al. 2020), and 5 from Apache Point Observatory (APO). These spectra are available on Zenodo (Hosseinzadeh et al. 2020). We used RVSAO (Kurtz & Mink 1998) to cross-correlate these spectra with a library of galaxy templates, including 6 built into RVSAO and an additional 10 from SDSS. Table 1 logs these observations and lists the modes of the resulting redshift distribution from RVSAO, ignoring matches with $z < 0.005$ ($d_L < 20$ Mpc).

We excluded from our analysis 1177 spectra where all template matches had a Tonry & Davis (1979) cross-correlation score $R_{CC} < 4$, meaning that the cross-correlation redshift may not be reliable, and 174 transients with only a single template match in the range $4 < R_{CC} < 5$. We then visually inspected approximately 600 host spectra that met one of three criteria: (1) the best (highest R_{CC}), median, and modal redshifts did not match; (2) no three redshift estimates matched each other; or (3) the best redshift could have been derived from matching a telluric feature or a known instrumental artifact to a feature in the galaxy template. In almost two-thirds of these cases, we were able to either verify the redshift from RVSAO or determine a new redshift manually. We excluded the remaining third from further analysis. In addition to the redshifts from RVSAO, we supplemented our sample with redshifts from publicly available catalogs[22] using a 1″ matching radius; in the case of a conflict with RVSAO, we manually inspected our spectrum to determine a final redshift. Lastly, two transient

[19] Documentation: https://griffin-h.github.io/superphot/.

[20] "Superphot" is a blend (Algeo 1977) of the words "supernova photometry" but is also intended to sound like the SN spectrum-fitting code Superfit (Howell et al. 2006) used for spectroscopic classification.

[21] Transients with three S/ N ⩾ 4 photometric observations in any filter and no history of variability (Jones et al. 2018).

[22] Dressler & Gunn (1992), Im et al. (2001), Colless et al. (2003), Szokoly et al. (2004), Le Fèvre et al. (2005), Cannon et al. (2006), Norris et al. (2006), Garcet et al. (2007), Lilly et al. (2007), Tajer et al. (2007), Bronder et al. (2008), Ross et al. (2008), Finkelstein et al. (2009), Jones et al. (2009), Lamareille et al. (2009), Owen & Morrison (2009), Scarlata et al. (2009), Trump et al. (2009), Balestra et al. (2010), Cowie et al. (2010), Drinkwater et al. (2010), Hewett & Wild (2010), Stalin et al. (2010), Rovilos et al. (2011), Cappellaro et al. (2012), Smith et al. (2012), Newman et al. (2013), Drout et al. (2014), Karhunen et al. (2014), Lunnan et al. (2014), Wen & Han (2015), Hasinger et al. (2018), Masters et al. (2019), Ahumada et al. (2020), Lidman et al. (2020).

Table 1
Host-galaxy Redshifts from RVSAO

Transient Name	Spec. Class.	Host R.A. (deg)	Host Decl. (deg)	MJD of Observation	Telescope	Final Redshift	RVSAO Redshift	Template Matches	Maximum R_{CC}
PSc000006	SN Ia	53.3663	−28.3715	57013	AAT	0.231	0.230	8	6.56
							0.109	2	2.40
PSc000010	SN Ia	149.7495	3.1576	57133	MMT	0.245	0.244	13	15.0
PSc000011	SN Ia	149.9760	2.4106	57013	AAT	0.380	0.731	3	6.77
							0.361	2	2.58
PSc000012	⋯	150.2308	1.8451	55296	MMT	0.623	0.179	3	3.12
							0.341	2	2.67
							0.727	2	1.89
PSc000013	⋯	149.1394	1.5447	56564	MMT	0.372	0.372	5	5.81
							0.927	2	3.60

(This table is available in its entirety in machine-readable form.)

spectra (of PSc110446 and PSc130816) yielded redshifts but not confident classifications; we treat these as unclassified transients with known redshifts.

Finally, we excluded 199 unclassified transients whose light curves are variable across multiple observing seasons, indicating that they are unlikely to be SNe. The remaining 2315 transients comprise our "test set" for photometric classification. Jones et al. (2017) previously used 1020 SNe Ia from this data set, some of which were photometrically classified by a different method, to constrain cosmological parameters.

3. Description of the Algorithm

3.1. Model Fitting

We use Equation (1) of V19 to model the single-band flux of a transient with the following form:

$$F(\Delta t) = \frac{A[1 - \beta \min(\Delta t, \gamma)]\exp\left(-\frac{\max(\Delta t, \gamma) - \gamma}{\tau_{\rm fall}}\right)}{1 + \exp\left(-\frac{\Delta t}{\tau_{\rm rise}}\right)}, \quad (1)$$

where $\Delta t \equiv t - t_0$ is time with respect to a reference epoch in the observer frame. This function has six parameters that are not strictly physical, but roughly correspond to an amplitude (A), the "plateau" slope and duration[23] (β and γ), the reference epoch with respect to discovery (t_0), and the (exponential) rise and decline times ($\tau_{\rm rise}$ and $\tau_{\rm fall}$).

To obtain not only the best-fit parameters but a quantification of the uncertainties, we use a Markov Chain Monte Carlo (MCMC) routine to fit each of the observed *griz* light curves for these six parameters plus an additional intrinsic scatter term, which is added in quadrature with the photometric uncertainties. The seven fit parameters and their priors are listed in Table 2, where $U(a, b)$ indicates a uniform distribution between a and b, $U_{\log}(a, b)$ indicates a log-uniform distribution between a and b, and $N(\mu, \sigma^2)$ indicates a Gaussian distribution with mean μ and variance σ^2. These are approximately the same priors used by V19, with the exception of A, for which they used a uniform prior. All parameters apart from t_0 are restricted from taking negative values.

[23] In the parameterization of V19, $\beta \rightarrow -\beta/A$ and $\gamma \equiv t_1 - t_0$.

Table 2
Model Parameters

	Parameter	Units	Prior
A	Amplitude	[flux]	$U_{\log}(1, 100F_{\rm max}^{\rm obs})$
β	Plateau Slope	days^{-1}	$U(0, 0.01)$
γ	Plateau Duration	days	$\frac{2}{3}N(5, 25) + \frac{1}{3}N(60, 900)$
t_0	Reference Epoch	days	$U(-50, 300)$
$\tau_{\rm rise}$	Rise Time	days	$U(0.01, 50)$
$\tau_{\rm fall}$	Fall Time	days	$U(1, 300)$
⋯	Intrinsic Scatter	[flux]	$N(0, 1)$

Because not all filters (*griz*) are well sampled for all events, we wish to include some cross-filter information in the fit. However, prior to classification, we cannot assume knowledge of the shape and evolution of the spectral energy distribution (SED). We also found it difficult to achieve MCMC convergence within a reasonable time when simultaneously fitting more than 20 parameters. We therefore adopt V19's two-iteration fitting approach that consists of fitting each filter separately, adding the posteriors from those fits together, and using the result as the prior for a second iteration of fitting (see Figure 1). This effectively weights the filters toward being more similar to each other, but without excluding the possibility that they are different. If a given SN was not observed in one or more filters, we average the posterior distributions for all the observed filters and treat that as the posterior for all unobserved filters.

For all fits, we use the Metropolis–Hastings sampling algorithm (Metropolis et al. 1953; Hastings 1970) in the PyMC3 package (Salvatier et al. 2016). We use 25 chains (walkers), each drawing 25,000 samples for tuning (burn-in) plus 10,000 samples for the posterior. For computational efficiency, we only fit points between −50 and +180 days of the discovery date, which encompasses >99% of the $\geqslant 3\sigma$ detections in the same observing season. The computations in this paper were run on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University.[24] Tables 3–5 list the reduced χ^2

[24] The Cannon cluster is named after Annie Jump Cannon, one of the human computers at the Harvard College Observatory and a pioneer in stellar classification.

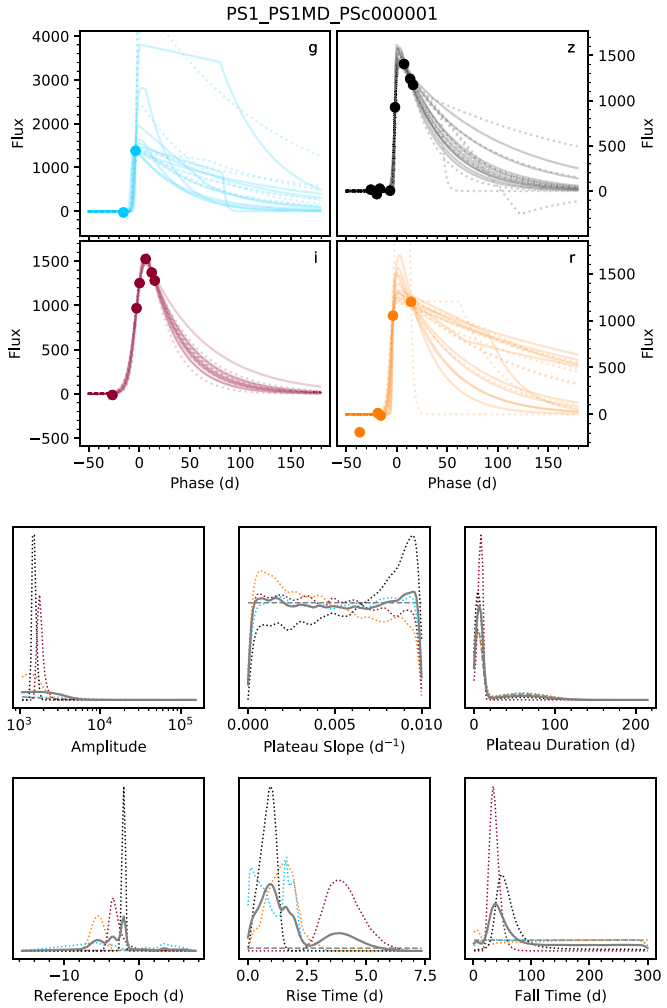

Figure 1. Two-step MCMC fit of Equation (1) to a sample transient. The top panels show the light-curve models in each of the four filters. Dotted lines represent the results of the first iteration, and solid lines represent the results of the second. The bottom panels show how the first-iteration posteriors are combined to produce the second-iteration priors. The gray dashed line is the first-iteration prior, the colored dotted lines represent the first-iteration posteriors for the four filters (same colors as the upper panels), and the gray solid line is the second-iteration prior.

goodness-of-fit statistic (for the model with the median parameters) and the maximum (among the 28 model parameters) Gelman & Rubin (1992) convergence statistic for each transient. In most cases, both statistics are close to their optimal value of 1. Keep in mind that the reduced χ^2 is infinite or negative when the light curve has 24 or fewer points across the four filters.

The resulting parameters are shown in Figure 2. As expected, the parameters for each photometric class in the test set (small markers) overlap with the parameters for the equivalent spectroscopic class in the training set (large markers). However, there is also significant overlap between different classes in this parameter space. In particular, SNe Ibc have peak magnitudes and evolution timescales similar to some SNe II, although SNe Ibc typically do not have a long plateau. Some SNe Ia also have long fall times, either because the declining light curve was not well constrained or because the secondary infrared peak was modeled as a smooth decline. We discuss this further in Section 4.1.

3.2. Feature Extraction

V19 explored four methods of feature extraction from the model light curves: (1) directly using the model parameters (plus the peak absolute magnitude) as features, (2) hand-selecting features based on the model light curves, (3) performing a principal component analysis (PCA) on the model light curves and using the PCA coefficients (plus the peak absolute magnitude) as features, and (4) using the downsampled model light curves themselves as features. Among their 24 pipelines, there was no clear trend for which of these methods was best. However, as their best-performing pipeline used the PCA method, we adopt that here. Our code also gives the option of using the model parameters directly, but we find that this gives slightly worse results (see Appendix D).

For the training set, we generate one model light curve for each SN from the median of the posterior parameter distributions. For the validation and test sets, we generate 10 model light curves for each SN by randomly drawing from the posteriors; these uncertainties will be accounted for in our classification probabilities later. We then convert the model fluxes to luminosities. This requires that we know the extinction $E(B-V)$ and redshift z for each transient, where the latter can be measured from the spectrum of either the transient or its host galaxy. In particular, we use luminosity distances calculated with the cosmological parameters of the Planck Collaboration et al. (2016), the Milky Way extinction maps of Schlafly & Finkbeiner (2011), the extinction law of Fitzpatrick (1999) with $R_V = 3.1$, and a cosmological K-correction factor of $1+z$ for all filters.[25] We do not consider host-galaxy extinction, because we have no way of estimating it for transients in the test set. V19 found that correcting to rest-frame times gave worse results, so we leave all times in the observer frame.

For each filter, we then perform a PCA on the model light curves (in luminosity) in our training set. For the purposes of the PCA, we evaluate the model at 1000 phases $0 \leqslant \Delta t \leqslant 300$ days. Importantly, we only include SNe from the training set when calculating the principal components because our goal is to produce a self-contained classification pipeline that can be applied to any new light curves. The light curves in the test set also tend to be sampled worse than those in the training set, likely because brighter targets were prioritized for spectroscopic follow-up. We then project the light curves in our test set onto the same principal components.

We use the first six PCA coefficients (Figure 3, top), which together explain >99.9% of the sample variance in each filter, plus the peak absolute magnitude (from the model light curve), as features for each single-filter light curve. These features do not have a simple physical interpretation, although coefficients on the first principal component (a generic rising and declining light curve) are strongly correlated with peak magnitude (see Appendix C). Light curves with a smooth exponential rise and decline (Figure 3, center) are reconstructed nearly perfectly from the principal components, whereas light curves with a plateau (Figure 3, bottom) are not reconstructed as well. Regardless, we find that the PCA coefficients are useful features for classification; a perfect reconstruction of the model light curve is not necessary.

[25] Correcting to standard rest-frame filters would require detailed knowledge of the SED and time evolution of each transient, which would in turn depend on its classification.

Table 3
Classification Results

Transient Name	Host-galaxy Redshift	Milky Way $E(B-V)$	Photometric Classification	Classification Confidence	Reduced χ^2	Maximum $\hat{R}$	Classification Probabilities				
							SLSN	SN II	SN IIn	SN Ia	SN Ibc
PSc000012	0.6226	0.0181	SN Ia	0.858	1.1	1.0	0.001	0.023	0.089	0.858	0.029
PSc000013	0.3720	0.0216	SN Ibc	0.843	1.2	1.0	0.000	0.002	0.024	0.131	0.843
PSc000015	0.2090	0.0286	SN II	0.574	17.5	1.0	0.000	0.574	0.003	0.077	0.346
PSc000017	0.2570	0.0213	SN II	0.757	2.8	1.0	0.000	0.757	0.023	0.040	0.180
PSc000022	0.2530	0.0244	SN II	0.665	2.5	1.1	0.000	0.665	0.015	0.062	0.258
PSc000031	0.2190	0.0250	SN Ia	0.592	2.9	1.0	0.002	0.024	0.375	0.592	0.007
PSc000032	0.1650	0.0298	SN II	0.566	3.3	1.0	0.000	0.566	0.002	0.079	0.353
PSc000036	2.0260	0.0196	SLSN	0.891	1.2	1.0	0.891	0.000	0.009	0.100	0.000
PSc000051	0.1940	0.0123	SN II	0.838	−3.2	1.1	0.000	0.838	0.010	0.042	0.110
PSc000059	0.7800	0.0245	SN Ia	0.384	35.5	1.0	0.376	0.006	0.224	0.384	0.010
PSc000060	0.1470	0.0270	SN II	0.560	1.8	1.0	0.000	0.560	0.000	0.144	0.296
PSc000068	0.1950	0.0256	SN II	0.895	2.3	1.0	0.000	0.895	0.038	0.026	0.041
PSc000069	0.3360	0.0274	SN Ia	0.997	1.2	1.1	0.000	0.000	0.001	0.997	0.002
PSc000070	0.2030	0.0260	SN Ibc	0.480	1.1	1.0	0.000	0.456	0.001	0.063	0.480
PSc000075	0.0820	0.0263	SN Ibc	0.483	4.3	1.0	0.000	0.455	0.000	0.062	0.483
PSc000080	0.4510	0.0283	SN Ia	0.986	1.4	1.0	0.000	0.005	0.002	0.986	0.007
PSc000095	0.3300	0.0289	SN Ia	0.983	1.1	1.0	0.000	0.000	0.003	0.983	0.014
PSc000102	0.2390	0.0143	SN Ia	0.381	801.5	1.7	0.016	0.144	0.357	0.381	0.102
PSc000150	0.2060	0.0092	SN Ia	0.985	2.4	1.8	0.000	0.011	0.002	0.985	0.002

(This table is available in its entirety in machine-readable form.)

In total, each multiband SN light curve has 28 features (Figure 4). Before classification, we rescale each feature to have zero mean and unit variance in the training set and apply the same scaling to the test set.

3.3. Classification

The differences in the observed rates of our five SN classes mean that our training set is unbalanced: 404 Type Ia, 93 Type II, 24 Type IIn, 19 Type Ibc, and 17 SLSNe. For our classifier to perform effectively on the minority classes, and to increase our sample size in general, we augment each class with additional feature sets by oversampling our training set. V19 explored two oversampling methods: the synthetic minority oversampling technique (SMOTE; Chawla et al. 2002) and multivariate-Gaussian (MVG) oversampling. We implement both options in our code but use the latter for our final classifications, oversampling all classes to have 1000 members. In agreement with V19, we find that MVG oversampling gives better results because it allows for features outside the original distribution. We use the implementation of SMOTE in the imbalanced-learn package (Lemaître et al. 2017) and implement our own imbalanced-learn-compatible MVG oversampler based on the `multivariate_normal` function in NumPy (Oliphant 2006).

V19 tested three supervised machine-learning algorithms for classification: a random forest, a support vector machine, and a multilayer perceptron (a type of neural network). Our code includes implementations of all three algorithms from the scikit-learn package (Pedregosa et al. 2011), but our final classifications use the random forest option with 100 decision trees, an entropy split criterion, and a maximum of five features, as this set of hyperparameters performed best for V19 (see also Appendix D). After training the random forest on the oversampled training set, we apply it to the 10 sets of features (from the 10 random posterior draws) for each transient in the test set to get 10 sets of classification probabilities. We then average the 10 sets of classification probabilities for each transient and adopt the classification with the highest probability.

4. Results and Validation

Applying our classification pipeline to the test set described in Section 2 yields 1257 photometrically classified SNe Ia, 521 SNe II, 298 SNe Ibc, 181 SNe IIn, and 58 SLSNe.[26] Table 3 gives the full list of classifications, and Figure 5 shows a sample of photometrically classified light curves. These are among the largest samples of each of these classes of SNe in the literature from a single survey. In the remainder of this section, we assess the performance of our algorithm in general and discuss how to use these classifications in practice.

4.1. Cross-validation

We first validate our classifier using leave-one-out cross-validation with the SNe in our training set. For each iteration of the cross-validation, we retrain the classifier on all but one of the SNe in our training set (still using the median parameters) and then use it to classify 10 sets of features derived from the posterior parameter distributions for that SN. We then average the 10 sets of probabilities to determine the cross-validation classifications listed in Table 4. Figure 6 shows the resulting confusion matrices.

In general our code performs well, with an overall classification accuracy of 82%. (See Appendix A for a glossary of terms.) This is dominated by the photometrically homogeneous SNe Ia,

[26] To exactly reproduce our results, a seed of 0 must be used in the pseudorandom number generators during feature extraction, oversampling, and classification.

Table 4
Cross-validation Results

Transient Name	Transient Redshift	Milky Way $E(B-V)$	Spectroscopic Classification	Photometric Classification	Classification Confidence	Reduced χ^2	Maximum $\hat{R}$	Classification Probabilities				
								SLSN	SN II	SN IIn	SN Ia	SN Ibc
PS0909006	0.2840	0.0426	SN Ia	SN Ia	0.841	∞	1.7	0.004	0.017	0.092	0.841	0.046
PS0909010	0.2700	0.0256	SN Ia	SN Ia	0.964	2.1	1.0	0.003	0.000	0.033	0.964	0.000
PS0910016	0.2300	0.0219	SN Ia	SN Ia	0.944	2.0	1.0	0.000	0.002	0.004	0.944	0.050
PS0910017	0.3200	0.0221	SN Ia	SN Ia	0.984	2.1	1.0	0.000	0.002	0.008	0.984	0.006
PS0910018	0.2650	0.0242	SN Ia	SN Ia	0.910	3.2	1.0	0.000	0.006	0.016	0.910	0.068
PS0910020	0.2420	0.0130	SN Ia	SN Ia	0.799	1.4	1.0	0.000	0.052	0.054	0.799	0.095
PS0910021	0.2560	0.0081	SN Ia	SN Ia	0.845	1.3	1.0	0.000	0.013	0.142	0.845	0.000
PSc000001	0.0710	0.0090	SN II	SN II	0.734	−12.9	1.7	0.000	0.734	0.030	0.095	0.141
PSc000006	0.2308	0.0083	SN Ia	SN Ia	0.469	16.9	1.0	0.000	0.061	0.018	0.469	0.452
PSc000010	0.2447	0.0224	SN Ia	SN Ia	0.942	2.6	1.0	0.000	0.008	0.001	0.942	0.049
PSc000011	0.3800	0.0177	SN Ia	SN Ibc	0.683	1.0	1.0	0.000	0.000	0.007	0.310	0.683
PSc000014	0.1369	0.0261	SN Ia	SN Ia	0.883	8.6	1.1	0.000	0.014	0.077	0.883	0.026
PSc000034	0.2500	0.0278	SN Ia	SN Ia	0.706	1.3	1.0	0.011	0.021	0.244	0.706	0.018
PSc000038	0.1500	0.0220	SN Ia	SN Ia	0.996	1.6	1.0	0.000	0.001	0.001	0.996	0.002
PSc000076	0.2600	0.0245	SN II	SN II	0.757	1.1	1.0	0.000	0.757	0.004	0.049	0.190
PSc000091	0.1520	0.0269	SN Ia	SN Ia	0.599	3.1	2.9	0.000	0.005	0.003	0.599	0.393
PSc000098	0.0570	0.0127	SN II	SN II	0.673	4.8	3.0	0.000	0.673	0.009	0.049	0.269
PSc000133	0.2440	0.0082	SN II	SN Ia	0.373	3.4	1.0	0.000	0.285	0.183	0.373	0.159
PSc000137	0.1183	0.0080	SN Ia	SN Ia	0.780	2.8	1.0	0.000	0.005	0.003	0.780	0.212

(This table is available in its entirety in machine-readable form.)

Table 5
Rare Transients

Transient Name	Transient Redshift	Milky Way $E(B-V)$	Spectroscopic Classification	Photometric Classification	Classification Confidence	Reduced χ^2	Maximum $\hat{R}$	Classification Probabilities				
								SLSN	SN II	SN IIn	SN Ia	SN Ibc
PS0910012	0.0790	0.0073	SN Iax	SN II	0.830	4.8	3.0	0.000	0.830	0.002	0.056	0.112
PSc010411	0.0740	0.0091	FELT	SN II	0.515	6.2	1.0	0.000	0.515	0.000	0.097	0.388
PSc040777	0.1680	0.0134	TDE	SN IIn	0.772	1.2	1.0	0.000	0.036	0.772	0.186	0.006
PSc080333	1.3883	0.0537	Lensed SN Ia	SLSN	0.853	1.0	1.0	0.853	0.000	0.038	0.109	0.000
PSc091902	0.1120	0.0563	FELT	SN Ibc	0.477	1.4	1.0	0.000	0.322	0.003	0.198	0.477
PSc120170	0.4046	0.0303	TDE	SN IIn	0.637	2.4	1.0	0.023	0.015	0.637	0.323	0.002
PSc150020	0.3230	0.0191	FELT	SN Ia	0.741	1.5	1.0	0.046	0.000	0.140	0.741	0.073
PSc340012	0.6460	0.0302	FELT	SN Ia	0.704	1.3	1.0	0.070	0.001	0.186	0.704	0.039
PSc350224	0.1010	0.0300	FELT	SN Ibc	0.575	2.3	1.1	0.000	0.394	0.000	0.031	0.575
PSc350352	0.4050	0.0118	FELT	SN Ia	0.727	1.4	1.1	0.032	0.000	0.164	0.727	0.077
PSc370290	0.0535	0.0310	SN Ibn	SN Ia	0.567	6.0	2.2	0.000	0.060	0.159	0.567	0.214
PSc370330	0.1760	0.0193	SN IIb?	SN Ibc	0.779	0.9	1.0	0.000	0.046	0.037	0.138	0.779
PSc440088	0.2750	0.0970	FELT	SN Ia	0.692	1.8	1.0	0.003	0.022	0.066	0.692	0.217
PSc570006	0.2693	0.0638	FELT	SN Ia	0.682	2.3	1.1	0.027	0.013	0.173	0.682	0.105
PSc570060	0.2450	0.0620	FELT	SN II	0.813	3.1	1.0	0.000	0.813	0.012	0.055	0.120
PSc580304	0.2960	0.0292	FELT	SN II	0.400	3.5	1.0	0.000	0.400	0.089	0.385	0.126

(This table is available in machine-readable form.)

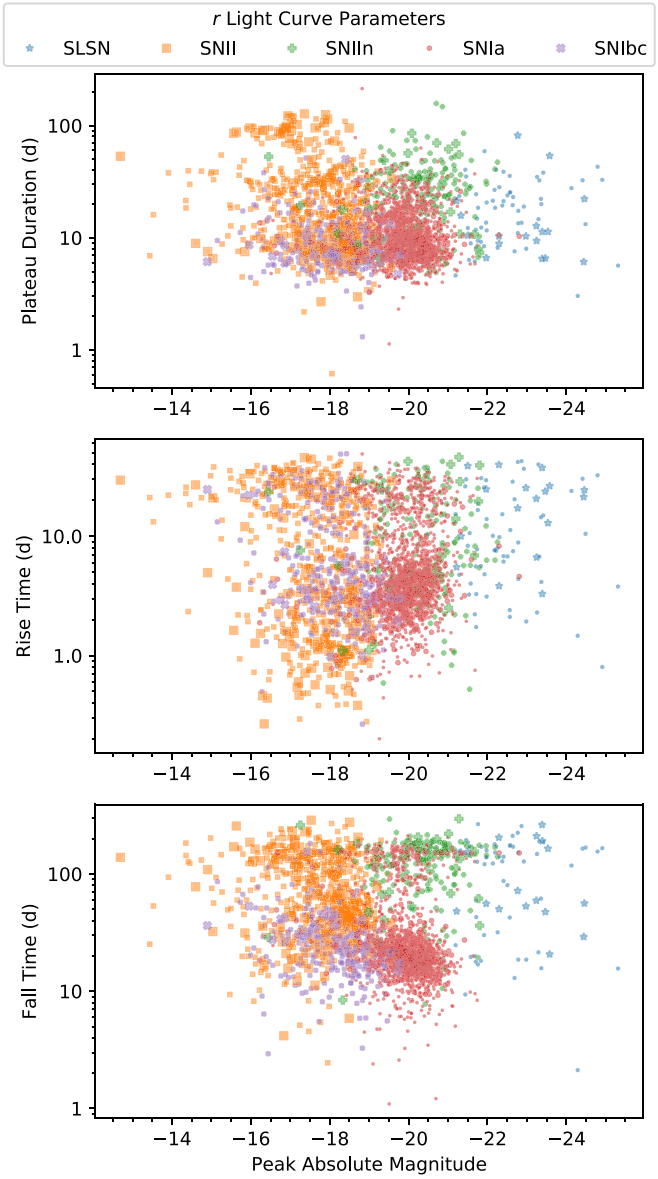

Figure 2. The plateau duration (top), rise time (center), and fall time (bottom) of each model *r*-band light curve plotted against its peak absolute magnitude. Large markers represent the training set and are colored by spectroscopic classification. Small markers represent the test set and are colored by photometric classification. The overlap between the SN classes in this parameter space demonstrates the challenge of photometric classification.

which are 85% complete. SLSNe, which are 82% complete, are also easy to identify because they separate relatively cleanly by absolute magnitude. SNe IIn (50% complete) and Ibc (37% complete) are hardest to identify because their light curves are intrinsically more heterogeneous, and because we have relatively small numbers of each in the training set, so their diversity is not well sampled. SNe Ibc also overlap significantly with SNe II in feature space (Figure 4).

In addition to being complete, most of our photometric classes are relatively pure: 96% for SNe Ia, 73% for SNe II, and 67% for SLSNe. For SNe IIn and Ibc, our purity falls to 39% and 21%, respectively. The latter raises an important point: in a magnitude-limited survey, where SNe Ia make about 70% of observed SNe, a small fraction of misclassified SNe Ia can significantly contaminate rarer classes. For example, in our

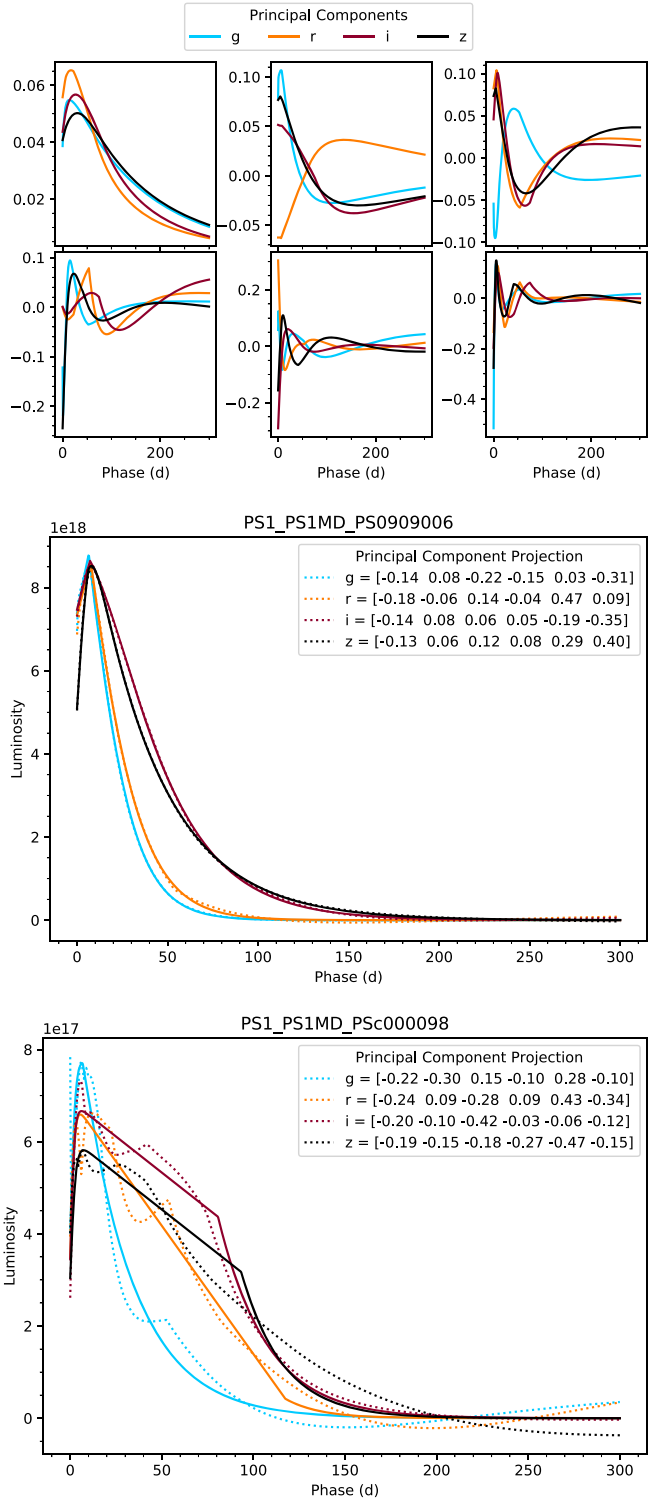

Figure 3. Top: the top six principal components for light curves in the training set. The projection of light curves onto these axes provides six of the features for each filter, the seventh being peak absolute magnitude. Center: an example of an SN Ia light curve (solid lines) that is reconstructed nearly perfectly (dotted lines) by its projection onto the principal components. Bottom: an example of an SN II light curve, which is not reconstructed as well. In both cases, these PCA coefficients were sufficient to classify the SN correctly (see Table 4).

data set, the 5% of SNe Ia misclassified as SNe Ibc represent 65% of our photometrically classified SNe Ibc.

The most common misclassification is labeling 42% of SNe Ibc as SNe II. Intriguingly, misclassifications in the

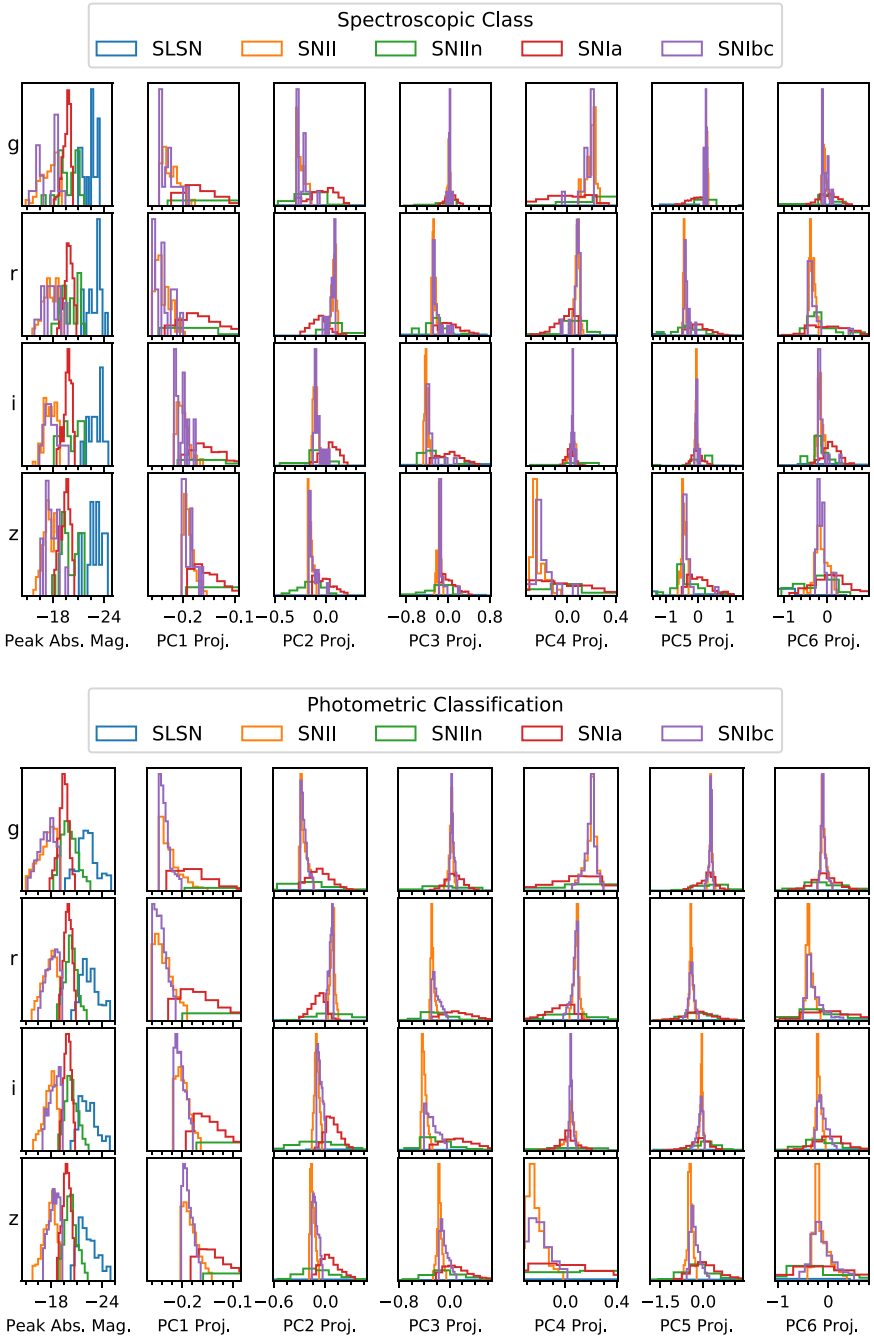

Figure 4. Histograms of the features (before rescaling) for each SN in the training (top) and test (bottom) sets, grouped by spectroscopic (top) and photometric (bottom) classification. Each row shows one filter. SN IIn and especially SLSN features extend outside the range of the axes. The classes do separate in this feature space. However, the significant overlap between SNe II and SNe Ibc, and to a lesser extent between SNe Ia and SNe Ibc, makes it difficult to identify members of the less numerous class (SNe Ibc).

opposite direction happen at much lower rates. This shows that the random forest has labeled a larger region of feature space as SN II (see also Figure 2). We suspect this is due to the fact that our SN II training set includes light curves with very flat plateaux as well as steep exponential declines (previously referred to as SNe IIP and IIL, respectively).

If we aggregate all classes other than SNe Ia under the label core-collapse SN (CCSN), this combined class is 95% complete and 66% pure (Figure 6, bottom). This shows that most of our misclassifications are between subtypes of CCSNe, rather than between CCSNe and SNe Ia. In the case of the binary classification, our sample of SNe Ia is actually purer (98% versus 94%), but less complete (81% versus 91%), than the final photometric sample of SNe Ia from the SDSS-II SN Survey (Sako et al. 2011).

4.2. Class Fractions

If we ignore any biases in selecting targets for spectroscopic follow-up,[27] the true class fractions in the photometrically classified sample should approximately match the class

[27] Our spectroscopic follow-up program serviced multiple science goals, so we consider the spectroscopic class fractions to be roughly representative of a magnitude-limited survey.

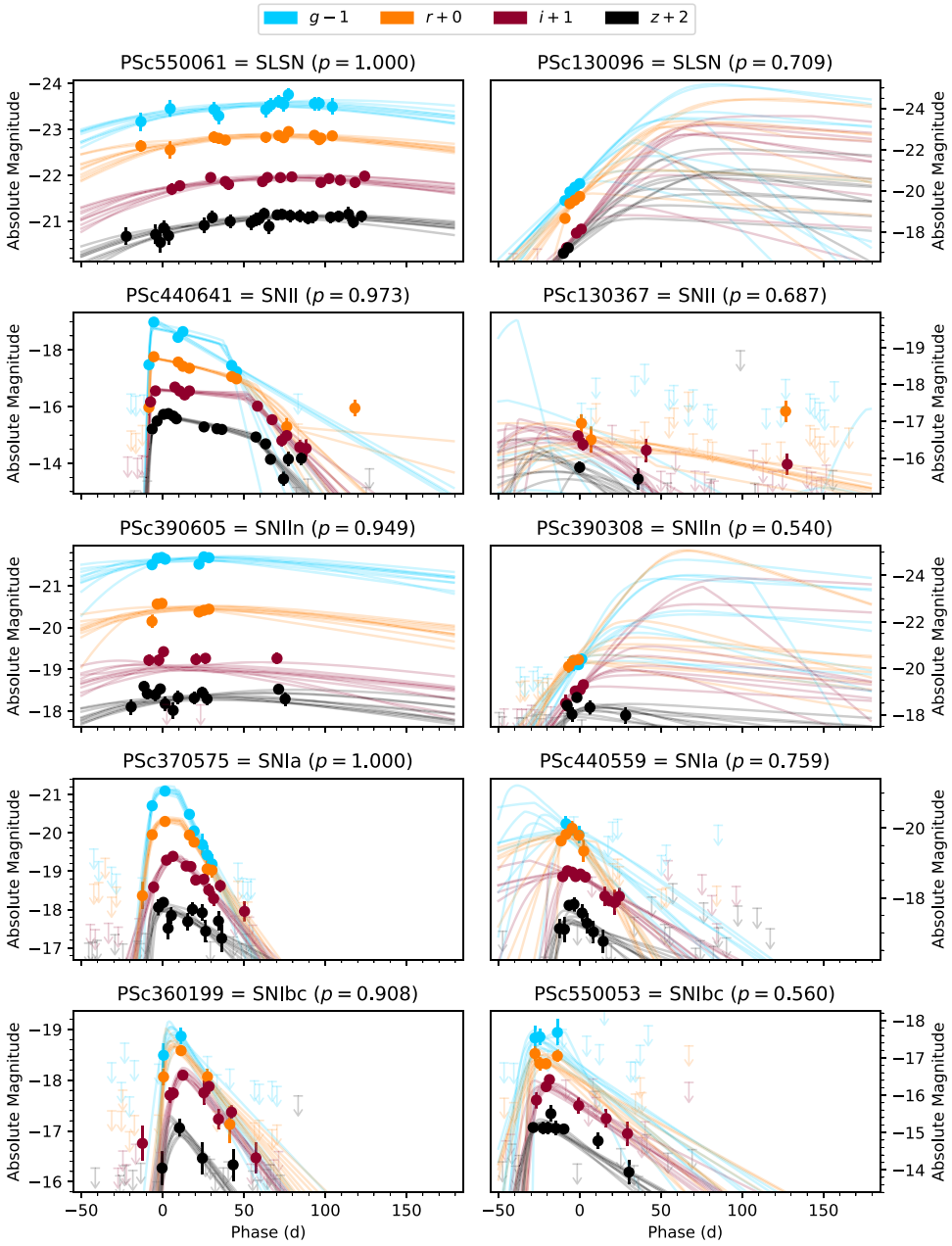

Figure 5. Sample observed (circles) and model (lines) light curves of photometrically classified SNe. Downward-pointing arrows indicate 3σ upper limits from nondetections. The left column shows the highest confidence classification in each class, and the right column shows the median-confidence classification. Notice that in three of the lower confidence classifications, the light curve is poorly sampled on either the rise or the decline because it was discovered close to the boundaries of the observing season.

fractions in the spectroscopically classified sample. However, our algorithm has different misclassification rates for each class of SNe, which we have measured using cross-validation. We can test the validity of these measured misclassification rates by using them to "correct" the class fractions in the photometrically classified sample and checking if the corrected fractions match the fractions in the spectroscopically classified sample.

For example, we observe that our photometrically classified sample contains a larger fraction of SNe II and Ibc and a smaller fraction of SNe Ia compared to our spectroscopically classified sample (Figure 7). From our confusion matrix (Figure 6, center), we can see that this is due to small fractions of SNe Ia contaminating the photometric SN II and Ibc samples. If we correct these fractions for the measured misclassification rates in our training set, we obtain a class breakdown similar to our training set (Figure 7, center). This suggests that the numbers in our confusion matrix are a good representation of the performance of our classifier.

4.3. Confidence Thresholds

We also assess how confident our classifier is in its predictions by examining the classification confidence (the highest classification probability) for each SN in the training set. Figure 8 shows cumulative histograms of the classification confidences for each spectroscopic class (top left) and photometric class (center left). As expected, we find that on average the classifier is most confident in predicting SLSNe

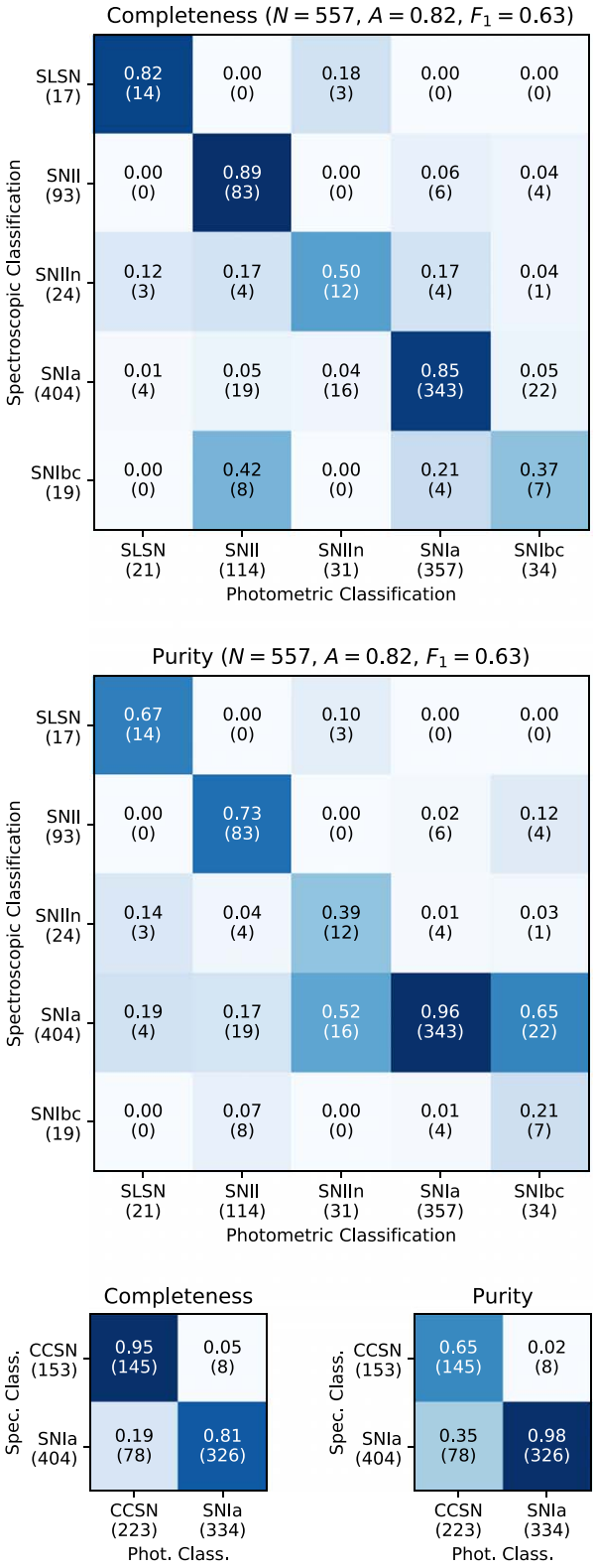

Figure 6. Confusion matrices of our validation results. Each cell lists and is colored by the fraction of each class, with the raw number in parentheses. The top matrix aggregates by true label, so its diagonal shows completeness. The middle matrix aggregates by predicted label, so its diagonal shows purity. The bottom matrices are the same results but with all four non-SN Ia classes grouped together as CCSNe. N is the total size of the training set, A is the accuracy, and the F_1 is a class-weighted average of completeness and purity (see Appendix A for definitions).

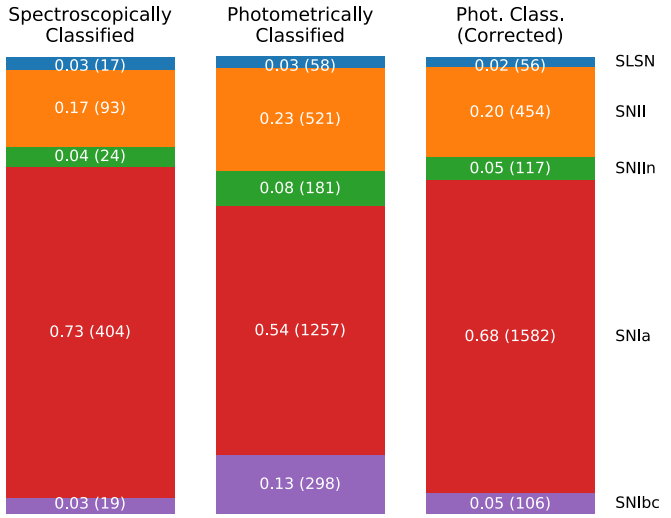

Figure 7. Fractions of our spectroscopic (left) and photometric (center) samples belonging to each class. The difference in class breakdown is likely due to small fractions of SNe Ia being misclassified as SNe II and Ibc. The "corrected" fractions (right) account for our expected misclassification rates (i.e., Figure 6). Under the assumption that we performed random spectroscopic follow-up, similarity between the composition of our spectroscopic and corrected photometric samples would suggest that we understand the performance of our classifier.

and SNe Ia, and least confident in SNe Ibc and IIn. In fact, 9 of the 12 highest confidence ($p > 0.8$) misclassifications are for spectroscopically classified SNe Ibc and IIn. We cannot assess the correctness of the classifications in the test set, but we observe that the distributions of the classification confidence in the test set (Figure 8, bottom left) are similar to those in the training set. This suggests that our claims about misclassification rates may generalize to the test set.

We can increase the completeness and/or purity of our photometrically classified samples by considering only transients classified with confidence above a certain threshold, at the expense of decreasing their absolute numbers. Figure 9 (top and center) shows how our performance metrics vary as a function of the confidence threshold chosen. There is no clear optimum for all classes, so any threshold is arbitrary. However, a threshold of $p \geqslant 0.75$, for example, is better than using the full sample for all classes but SNe Ibc, and only excludes 39% of the training set (48% of CCSNe). Figure 9 (bottom) shows the confusion matrix that results from imposing this threshold. With the exception of SNe Ibc (which are nearly eliminated), all classes are over 85% complete.

Likewise, we can improve our photometrically classified samples by requiring a certain number of photometric observations in order to remove poorly sampled light curves. Figure 8 shows analogous histograms of the number of $\geqslant 5\sigma$ detections (in all bands) for each spectroscopic (top right) and photometric (center right) class. We do find that misclassifications are more frequent for poorly sampled light curves, but the number of points required for a correct classification varies significantly between the spectroscopic classes. For example, most SNe IIn with fewer than 50 observations are misclassified, whereas most SNe Ia and SNe II require only 20–30 points. Light curves with fewer than 10 detections ($\sim$2 per filter) are almost always misclassified. This is close to the median number of detections in our test set (Figure 8, bottom

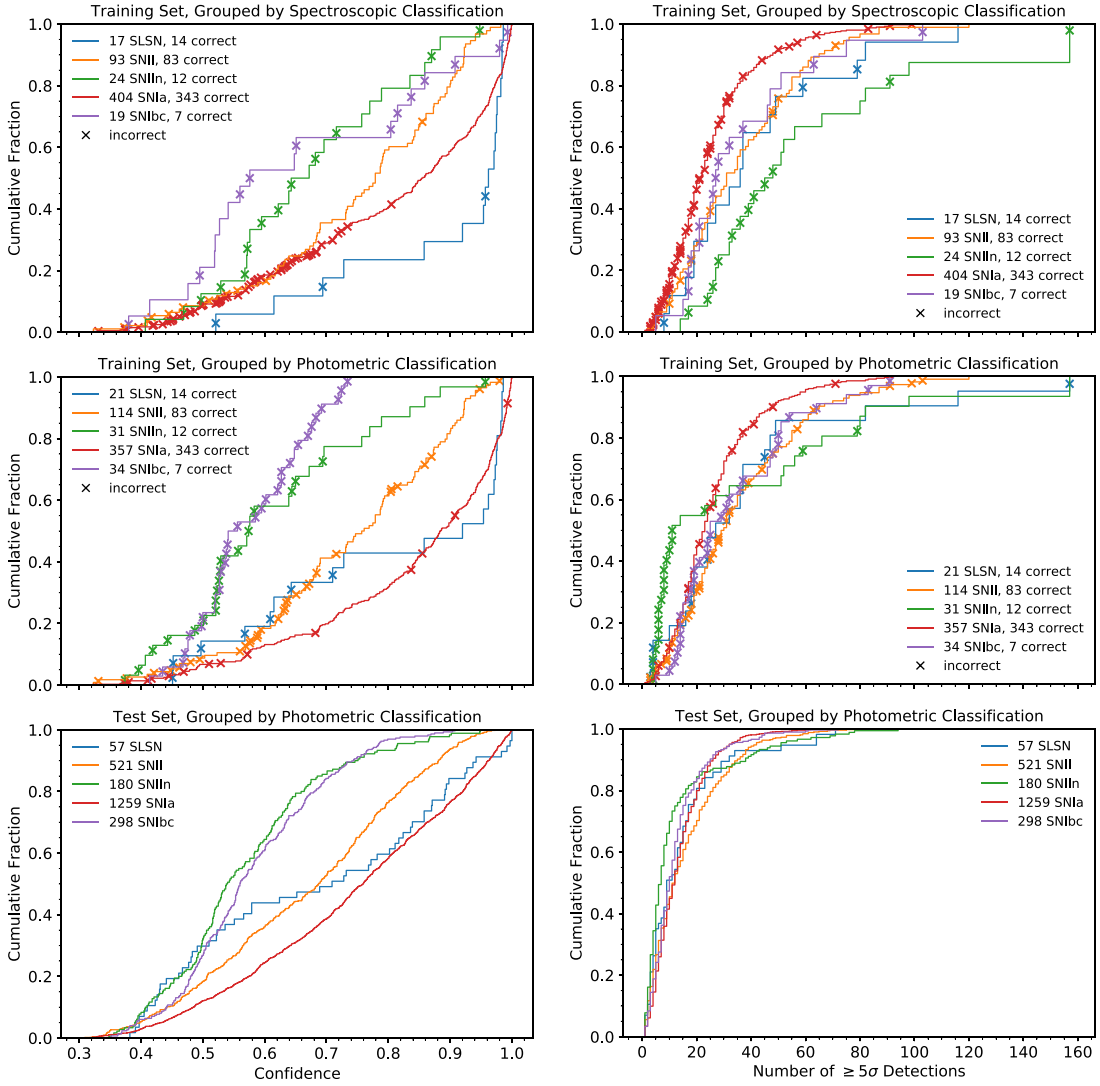

Figure 8. Cumulative histograms of the classification confidence (left column) and the number of $\geqslant 5\sigma$ detections (right column) for the training set (top and center rows) and the test set (bottom row), grouped by spectroscopic classification (top row) or photometric classification (center and bottom rows). Transients whose photometric and spectroscopic classifications do not match are marked by an $\times$. SLSNe and SNe Ia are typically classified with the highest confidence. With the exception of SNe Ibc, most false-positive classifications have low confidence and/or few detections.

right). Because of the clear difference in the distributions of the training and test sets, we do not know if a threshold chosen for the training set will have the desired effect on the test set. We also find that a confidence threshold is more effective and removes a smaller fraction of the sample, so we do not adopt a threshold on the number of detections.

4.4. Comparison to Other Photometric Classifiers

Although we cannot judge the correctness of individual classifications in our test set, we can see how often our algorithm agrees with other photometric classifiers applied to the same data set. In general, if two classifiers are independent for a given transient, we cannot expect their agreement matrix to be much better than the product of their confusion matrices (see Appendix B for a full derivation). Each has its own strengths and weaknesses, which can be assessed through validation.

In particular, we compare to the semisupervised machine-learning classifier of V20, SuperRAENN, which is somewhat less accurate than our classifier but has advantages in terms of speed and extensibility (see V20 for more details). We also compare to three CCSN versus SN Ia classifiers applied by Jones et al. (2017): the Photometric Supernova Identification (PSNID; Sako et al. 2011) code provided in the Supernova Analysis (SNANA; Kessler et al. 2009) package, which compares light curves to templates of SNe II, SNe Ia, and SNe Ibc; and the "Nearest Neighbor" and "Fitprob" classifiers, which compare light curves to the SALT2 SN Ia template (Jha et al. 2007) in parameter space and flux space, respectively. Figure 10 shows our "agreement matrices" with these classifiers. The numbers in the title of each panel indicate the number of transients we have in common (N) and the fraction of classifications we agree on (A'), and the diagonals show the fraction of our classifications that they agree with for each class.

As expected, V20 agree with 74% of our classifications overall, including large fractions of classes that we classify most accurately: 92% of SNe Ia, 65% of SNe II, and 48% of SLSNe. (Our agreement on cross-validation predictions on the training set, 83% overall, follows a similar pattern.) Figure 10 (top) reveals three important trends. First, V20 agree with 43%

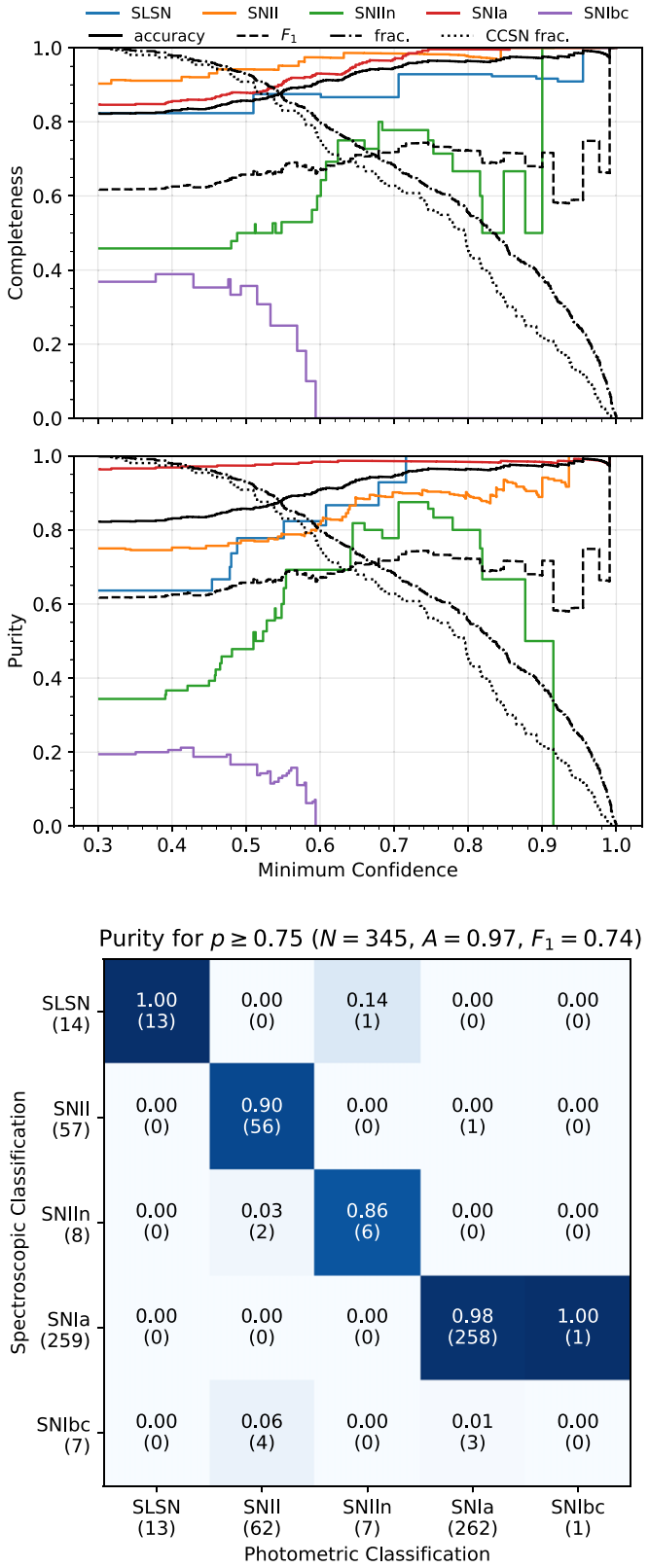

Figure 9. Completeness (top) and purity (center) for each class as a function of the minimum acceptable confidence. The total accuracy (solid), F_1 score (dashed), fraction of events remaining (dotted–dashed), and fraction of CCSNe remaining (dotted) are shown in black on both panels for reference. By only considering photometrically classified events above a certain threshold (e.g., $p \geqslant 0.75$, bottom), we can increase the purity of most of our samples, at the expense of decreasing the absolute number of events (in this case by 39%).

of our SNe Ibc, almost five times more than expected from from our cross-validation results. This implies that both of our classifiers tend to misclassify the same types of transients as SNe Ibc (i.e., they break the assumption of independence in Appendix B). Second, V20 tend to classify many more transients as SNe Ia than we do. This may be a result of the imbalanced training set, which their method cannot fully account for (see V20 for further discussion). Third, they classify 24 of our SLSNe (41%) as SNe Ia. This is surprising because both SLSNe and SNe Ia are relatively easy to identify photometrically.

A visual inspection of these light curves shows that most are missing either the rise or decline, due to the beginning or end of the observing season for that field. This appears to lead to a failure mode where some of the model light curves peak around SLSN luminosities, even if none of the observed photometry is that bright (see, e.g., Figure 5, top right). In a way, this is the desired behavior: if the peak is not observed, we want to include the possibility that the transient may peak at a flux brighter than the brightest observed data point. However, this introduces a bias toward higher luminosities, because the model light curves will never peak significantly below the brightest observed point. This is one reason we introduced a log-uniform prior on the amplitude, a modification to the method of V19.

Jones et al. (2017) also agree with 68% (Fitprob) to 83% (Nearest Neighbor) of our classifications, where we have again aggregated all classes other than SNe Ia under the label CCSN. However, they had a more specific goal than we do: to produce a pure sample of SNe Ia for the purpose of measuring cosmological parameters. Therefore, we expect their SN Ia samples to be less complete and their CCSN samples to be less pure than our samples. Our agreement matrix with PSNID, which Jones et al. (2017) adopt as their preferred classifier, can indeed be interpreted as reflecting their preference for SN Ia purity over CCSN purity.

When deciding between two conflicting classifications, one should take into account the purity of the samples produced by each classifier, as well as the relative rates of the two classes in a magnitude-limited survey. For example, for the 94 transients that we classify as SNe Ibc and V20 classify as SNe Ia, we prefer the SN Ia classification for the majority of them, because these are much more common in nature and we know that our SN Ibc purity is only 21%. Of course, this type of analysis also depends on one's science goals (e.g., purity or completeness of a sample).

5. Discussion

5.1. Future Applications

Much of the success of our algorithm (unlike, e.g., Jones et al. 2017, 2018) depends on the discriminating power of absolute magnitude, for which we need a redshift. While we still used spectroscopy to determine these redshifts, most of them were determined after the transients had faded by observing their host galaxies with a multifiber spectrograph, which is much less time consuming than classification spectroscopy of one transient at a time. However, this is still not scalable to the sample sizes expected from LSST. The performance of the algorithm has not yet been tested with photometric redshifts. Graham et al. (2018) suggest that LSST

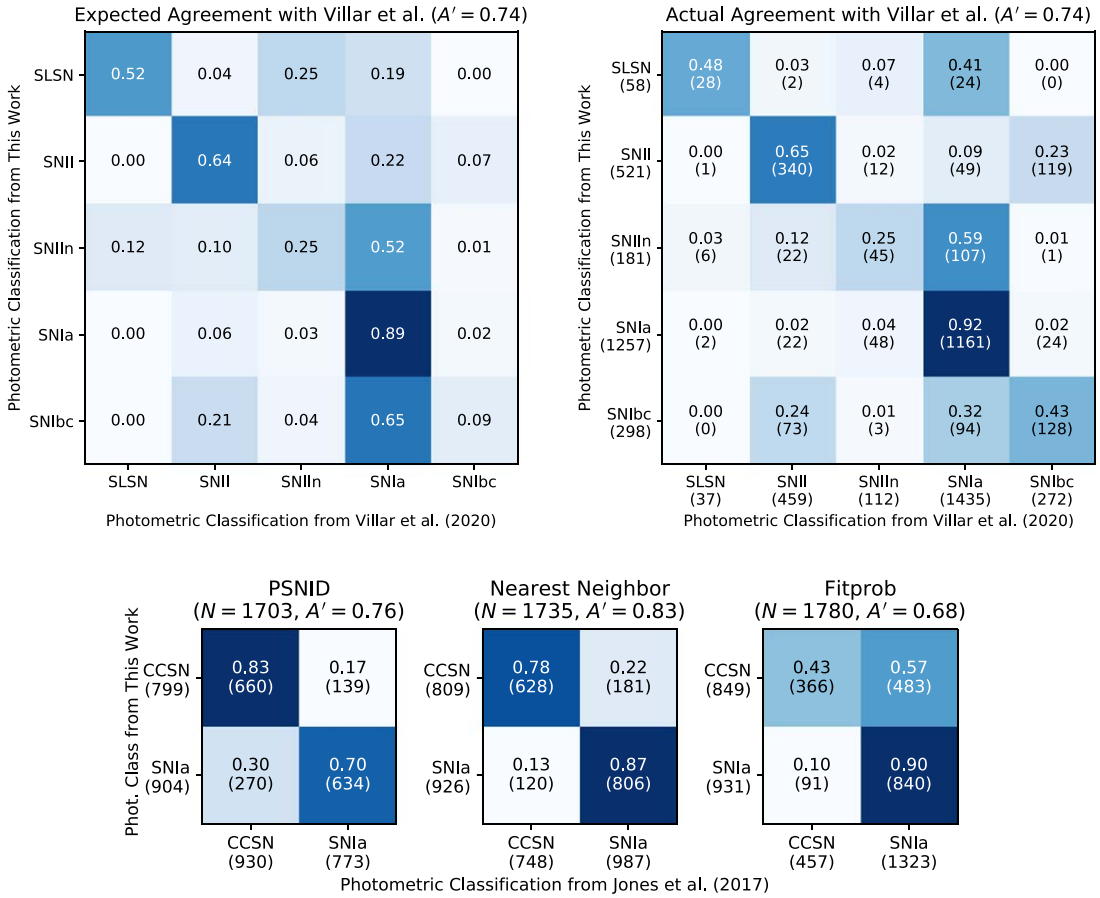

Figure 10. Top left: expected agreement between our classifier and that of V20, given our confusion matrices from validation (see Appendix B). We expect very low agreement on SNe IIn and Ibc, which both classifiers independently struggle with. Top right: actual agreement with V20. As expected, we agree on $A' = 74\%$ of classifications overall, including a larger than expected fraction of SNe Ibc (43%). Bottom: agreement between our classifier and those of Jones et al. (2017), where all classes that are not SNe Ia are aggregated under the CCSN label.

will determine photometric redshifts to ∼5% accuracy for galaxies with $r \lesssim 25$ mag within the first two years of survey operations. This would be only a small contribution to our classification uncertainties.

Furthermore, our classifications rely on the full light curves of these transients, in contrast to other algorithms that aim to classify transients in real time (e.g., Muthukrishna et al. 2019; Sravan et al. 2020). Our code will therefore be most successful at the end of an observing season, when the user has a large training sample of spectroscopically classified transients in hand and wants to make scientific use of the remaining transients with light curves only. However, our Bayesian light-curve modeling allows for fitting only part of the light curve while keeping all possible future behavior within the parameter uncertainties. Future work will explore what fraction of the light curve is required for good results.

Lastly, increasing the size of the training set would likely improve our results significantly, especially for SNe Ibc. With so few examples to train on, the algorithm is very sensitive to including or excluding even single events, as demonstrated by our low cross-validation scores for that classes (Figure 6). However this is not a shortcoming of the algorithm, but rather a reflection of the scarcity of large SN Ibc samples in the literature (Bianco et al. 2014; Taddia et al. 2015, 2018; Stritzinger et al. 2018).

Our current classifier returns probabilities determined entirely by the photometric data. In principle, we could multiply these probabilities by the relative rates of the various classes of SNe observed in previous surveys (e.g., Graur et al. 2017a, 2017b; Fremling et al. 2019; Holoien et al. 2019). This would have the effect of biasing borderline cases toward a more common classification; for example, it would decrease the number of transients that we classify as SLSNe and V20 classify as SNe Ia. We choose not to adopt such a prior in this work so that future analyses have the option of adopting the rate measurements of their choice. In practice, users may also want to combine photometric classification with contextual classification (e.g., Foley & Mandel 2013; Baldeschi et al. 2020; N. Chou et al. 2020, in preparation; Gomez et al. 2020).

5.2. Rare Classes of Transients

Sixteen of our spectroscopically classified transients did not belong to any of the five classes we consider here: one SN Iax (Narayan et al. 2011), two tidal disruption events (TDEs; Gezari et al. 2012; Chornock et al. 2014), one lensed SN Ia (Quimby et al. 2013), one SN Ibn (Sanders et al. 2013), one possible SN IIb (V20),[28] and ten fast-evolving luminous transients (FELTs; Drout et al. 2014). Presumably there are additional examples of these types of transients in our test set,

[28] Because SN IIb is a time-dependent classification (i.e., hydrogen features weaken during the evolution of the SN) and most of our classifications are determined by a single spectrum per transient, it is likely that other SNe IIb are "misclassified" as SNe II or SNe Ibc.

but because our classifier has no ability to identify them, they will contaminate our five photometric samples at a low level. By passing the light curves of known rare transients to the classifier, we can investigate how it might classify unknown rare transients.

Table 5 lists our classification results for these 16 transients. The SN Iax is classified as an SN II with high confidence, likely because of its low luminosity. Both TDEs are classified as SNe IIn, likely because of their high luminosity and slow evolution. The lensed SN Ia is classified as an SLSN with high confidence, likely because of its high luminosity. The SN Ibn is classified as an SN Ia with relatively low confidence, likely because it peaks at about the same luminosity as SNe Ia. The possible SN IIb is classified as an SN Ibc. The FELTs fall into all three classes that do not typically have slow evolution: SNe Ia, Ibc, and II. FELTs span a range of peak luminosities ($-16.5 > M > -20$ mag; Drout et al. 2014), and their fast evolution means that their light curves are not well sampled at the cadence of PS1-MDS. In general, we conclude that our classifier behaves as expected for these rare transients based on their peak luminosities and evolution timescales, but we note that in these cases a high confidence does not imply a correct classification.

5.3. Active Galactic Nuclei

In constructing the set of "SN-like" transients, we excluded light curves with a history of variability, with the intention of removing active galactic nuclei (AGNs) from the test set. However, if any AGNs survived this qualitative cut, they would likely be classified as SLSNe due to their high luminosities and slow evolution. To check for this possibility, we inspect the host-galaxy spectra of the photometric SLSNe to look for broad emission lines (a signature of accretion onto the central supermassive black hole). Not all of the spectra have a high enough signal-to-noise ratio to identify broad lines, but in at least 17 cases, they are visible. Of these, 14 of the transients are within $1''$ of the host center—PSc000478, PSc010120, PSc010186, PSc020026, PSc030013, PSc052281, PSc110163, PSc130394, PSc130732, PSc350614, PSc390545, PSc400050, PSc480585, PSc550061 (the latter is shown in Figure 5)—meaning that the AGN and the transient may be one and the same. Many (but not all) of these light curves are near the detection threshold, which could either indicate a nuclear SN (or even a TDE) that is faint compared to its AGN host, or a slight increase in the luminosity of the AGN itself. Because we cannot distinguish between these two cases, we urge caution in using these classification results.

6. Conclusions

We have presented the SN photometric classification package Superphot, based on the algorithm of V19. Training and then validating the classifier on 557 spectroscopically classified SNe from the Pan-STARRS1 PS1-MDS, we find that it has an overall accuracy of 82% and completenesses (purities) of 89% (73%) for SNe II, 85% (96%) for SNe Ia, 82% (67%) for SLSNe, 50% (39%) for SNe IIn, and 37% (21%) for SNe Ibc. We then apply this to 2315 previously unclassified transients from PS1-MDS for which we have robust host-galaxy redshifts, resulting in 1257 photometrically classified SNe Ia, 521 SNe II, 298 SNe Ibc, 181 SNe IIn, and 58 SLSNe.

In the process of validating our results, we raised several issues that will be relevant to future photometric classification efforts.

1. A small misclassification rate of SNe Ia can easily dominate photometric samples of minority classes like SNe Ibc.
2. SNe II overlap significantly with SNe Ibc in feature space, likely due to the subset of SNe II with linearly declining light curves.
3. Adopting a threshold on the classification confidence can improve the completeness and purity of the photometric samples, but an analogous threshold on the number of light-curve points is not as effective.
4. Agreement between two classifiers on a given transient is not necessarily an indication that they are correct; they may both be biased to misclassify certain transients in the same way.
5. Users should take into account the relative rates of different classes of SNe in addition to the photometric classification probabilities.
6. Transients belonging to none of the target classes can be misclassified into one of these classes with relatively high confidence.
7. AGNs may be a significant contaminant in photometrically classified SLSN samples.

Along with V20, this is the first application of a multiclass machine-learning classifier to a large photometric data set. As such, it serves as an example of the utility (and also the challenges) of photometric classification in the era of large time-domain surveys. Given that currently only a small fraction of transients discovered are classified spectroscopically, and the reality that this fraction will only decrease as discovery rates increase, we will have to increasingly rely on methods like this to extract as much science as possible from our data.

In addition, the photometric samples presented here are among the largest in the literature for each class, demonstrating the power of photometric classification to enable statistical studies of SNe. Importantly, however, each classification comes with an uncertainty. In the coming years, our field will have to learn how to handle exactly this type of photometric data set, when we will never know with certainty whether an individual classification is "correct." No single classifier will likely outperform the others for all use cases, but continued testing of algorithms individually and in combination will demonstrate how best to apply them toward a specific science goal.

We thank Jessica Mink and Brian Hsu for assisting with the host-galaxy redshifts. We also thank the authors of the "Scientific Python Cookiecutter" tutorial for advice on how to document, package, and release the Superphot package. The Berger Time-Domain Group is supported in part by NSF grant AST-1714498 and NASA grant NNX15AE50G. We acknowledge partial funding support from the Harvard Data Science Initiative. G.H. thanks the LSSTC Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining grant #1829740, the Brinson Foundation, and the Moore Foundation; his participation in the program has benefited this work. F.D. thanks the SAO REU program, funded in part by the National Science Foundation REU and Department of Defense ASSURE programs under NSF grant No. AST-1852268 and by the Smithsonian Institution. V.A.V. acknowledges support by the Ford Foundation through a

Dissertation Fellowship and the Simons Foundation through a Simons Junior Fellowship (#718240). D.O.J. is supported by a Gordon and Betty Moore Foundation postdoctoral fellowship at the University of California, Santa Cruz. The UCSC team is supported in part by NASA grants 14-WPS14-0048, NNG16PJ34C, NNG17PX03C; NSF grants AST-1518052 and AST-1815935; NASA through grant No. AR-14296 from the Space Telescope Science Institute, which is operated by AURA, Inc., under NASA contract NAS 5-26555; the Gordon & Betty Moore Foundation; the Heising-Simons Foundation; and by fellowships from the Alfred P. Sloan Foundation and the David and Lucile Packard Foundation to R.J.F. R.L. is supported by a Marie Skłodowska-Curie Individual Fellowship within the Horizon 2020 European Union (EU) Framework Programme for Research and Innovation (H2020-MSCA-IF-2017-794467). D. M. acknowledges NSF support from grants PHY-1914448 and AST-2037297.

The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg, and the Max Planck Institute for Extraterrestrial Physics, Garching, Johns Hopkins University, Durham University, the University of Edinburgh, Queen's University Belfast, the Center for Astrophysics | Harvard & Smithsonian, Las Cumbres Observatory, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation grant No. AST-1238877, the University of Maryland, Eötvös Loránd University (ELTE), Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

Facilities: ADS, NED, PS1.

Software: ArviZ (Kumar et al. 2019), Astropy (Astropy Collaboration et al. 2018), extinction (Barbary 2016), imbalanced-learn (Lemaître et al. 2017), IPython (Perez & Granger 2007), Matplotlib (Hunter 2007), NumPy (Oliphant 2006), PyMC3 (Salvatier et al. 2016), RVSAO (Kurtz & Mink 1998), scikit-learn (Pedregosa et al. 2011), SciPy (Virtanen et al. 2020), Theano (Theano Development Team 2016), `tqdm` (da Costa-Luis 2019).

Appendix A
Glossary

For each SN, the classifier gives a set of five classification probabilities (p_Φ) corresponding to each of the five photometric classes. They sum to 1 and are listed in Tables 3–5:

$$p_\Phi \in \{p_{\rm SLSN}, p_{\rm SNII}, p_{\rm SNIIn}, p_{\rm SNIa}, p_{\rm SNIbc}\}; \quad \sum_\Phi p_\Phi = 1.$$

An SN's photometric classification (Φ) is determined by its highest classification probability. We take this probability to be the classification confidence (p):

$$p \equiv \max\{p_\Phi\}.$$

$N_{S\Phi}$ is the number of SNe with spectroscopic classification S and photometric classification Φ; these are the elements of the confusion matrix (the integers in Figure 6). $N_S \equiv \sum_\Phi N_{S\Phi}$ is the total number of SNe with spectroscopic classification S, $N_\Phi \equiv \sum_S N_{S\Phi}$ is the total number of SNe with photometric classification Φ, and $N \equiv \sum_S \sum_\Phi N_{S\Phi}$ is the total sample size. $N_{S=\Phi}$ is the number of correctly classified SNe in a given class.

We discuss four performance metrics for our classifier, all of which range from 0 to 1:

1. Completeness (C_S) is the fraction of a given spectroscopic class that appears in the equivalent photometric class:
$$C_S \equiv \frac{N_{S=\Phi}}{N_S}.$$
2. Purity (P_Φ) is the fraction of a given photometric class that belongs to the equivalent spectroscopic class:
$$P_\Phi \equiv \frac{N_{S=\Phi}}{N_\Phi}.$$
3. Accuracy (A) is the total fraction of correctly classified SNe:
$$A \equiv \frac{\sum_{S=\Phi} N_{S\Phi}}{N}.$$
A variant of accuracy is agreement (A'), in which we compare two photometric classifications to each other, rather than comparing a photometric classification to a spectroscopic classification.
4. The (macro-averaged) F_1 score is the average of the harmonic means of the completeness and purity of each class:
$$F_1 \equiv \frac{1}{5}\sum_{S=\Phi}\frac{2}{C_S^{-1}+P_\Phi^{-1}} = \frac{1}{5}\sum_{S=\Phi}\frac{2N_{S\Phi}}{N_S+N_\Phi}.$$

Appendix B
Agreement between Two Classifiers

In Section 4.4, we compared our photometric classifications to those of V20. Here we derive the expectation for such a comparison given the confusion matrices for each classifier calculated from cross-validation.

The version of the confusion matrix with completeness on the diagonal (Figure 6, top) shows the probability $p(\Phi|S)$ of our classifier giving the photometric classification Φ for an SN with spectroscopic classification S. In this section, we will refer to this as the "completeness matrix" $\boldsymbol{C}$. The version of the confusion matrix with purity on the diagonal (Figure 6, center) shows the probability $p(S|\Phi)$ of an SN having spectroscopic classification S if we gave it a photometric classification Φ. In this section, we will refer to this as the "purity matrix" $\boldsymbol{P}$.

The "agreement matrix" $\boldsymbol{A}$ we wish to derive will show the probability $p(\Phi'|\Phi)$ of another classifier giving photometric classification Φ' to a transient that we classify as Φ. Using the chain rule of probability, we can write

$$p(\Phi'|\Phi) = \frac{p(\Phi, \Phi')}{p(\Phi)}.$$

The joint probability $p(\Phi, \Phi')$ cannot be separated because the classifiers are not independent. (If they were, they would not be good classifiers.) However, we assume that the classifiers are independent for a given spectroscopic class, meaning that, if they are biased, they are not biased in the same way. We can then obtain a separable joint probability by undoing the

marginalization over spectroscopic classification:

$$p(\Phi'|\Phi) = \sum_S \frac{p(\Phi, \Phi'|S)p(S)}{p(\Phi)} = \sum_S \frac{p(\Phi|S)p(\Phi'|S)p(S)}{p(\Phi)}.$$

Lastly, we simplify using Bayes's theorem (Bayes & Price 1763):

$$p(\Phi'|\Phi) = \sum_S p(S|\Phi)p(\Phi'|S),$$

or in matrix notation,

$$\boldsymbol{A} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{C}',$$

where $\boldsymbol{C}'$ is the completeness matrix of the other classifier. Note that this matrix depends on the breakdown of spectroscopic classes in the data sets used to validate these classifiers. Figure 10 (top left) shows the expected agreement matrix between our classifier and that of V20, assuming that our test set has similar class fractions to the training set.

Appendix C
Feature Importances

Of interest to developers of future photometric classification algorithms is the question of which features are most powerful for classifying SNe. V19 explore several combinations of features, and we adopt their best-performing classifier, which uses peak absolute magnitude in the *griz* filters plus coefficients of the top six principal components of the *griz* light curves. Here we explore the relative importances of each of those features.

The importance of a feature cannot be defined independently of the other features used. For example, if two features are perfectly correlated, a classifier may arbitrarily consider one of them to be very important and the other to be useless. Therefore, before calculating feature importances, we must examine correlations between the features in our set. Figure 11 shows the absolute value of the Spearman (1904) rank correlation coefficient between every pair of features (top) and model parameters (bottom) in our training set.

As we might expect, there are strong correlations between the four filters, both because physics demands a relatively smooth SED and because our two-iteration fitting method forces the light-curve models in the four filters to be more similar. This means it is not possible to judge whether, for example, the *g* peak absolute magnitude is more important than the *r* peak absolute magnitude. In addition, we find that the peak absolute magnitudes are strongly correlated with the light curves' projection onto their principal component. This tells us that most of the variation among the light curves can be attributed to differences in overall luminosity, rather than differences in shape.

To obtain meaningful feature importances, we retrained our classifier on only one filter (seven features) at a time. We then calculated two measures of feature importance for each of the seven features: mean decrease in impurity (Louppe 2015) and permutation importance (Breiman 2001). We also calculate the permutation importance of a random feature, for comparison,

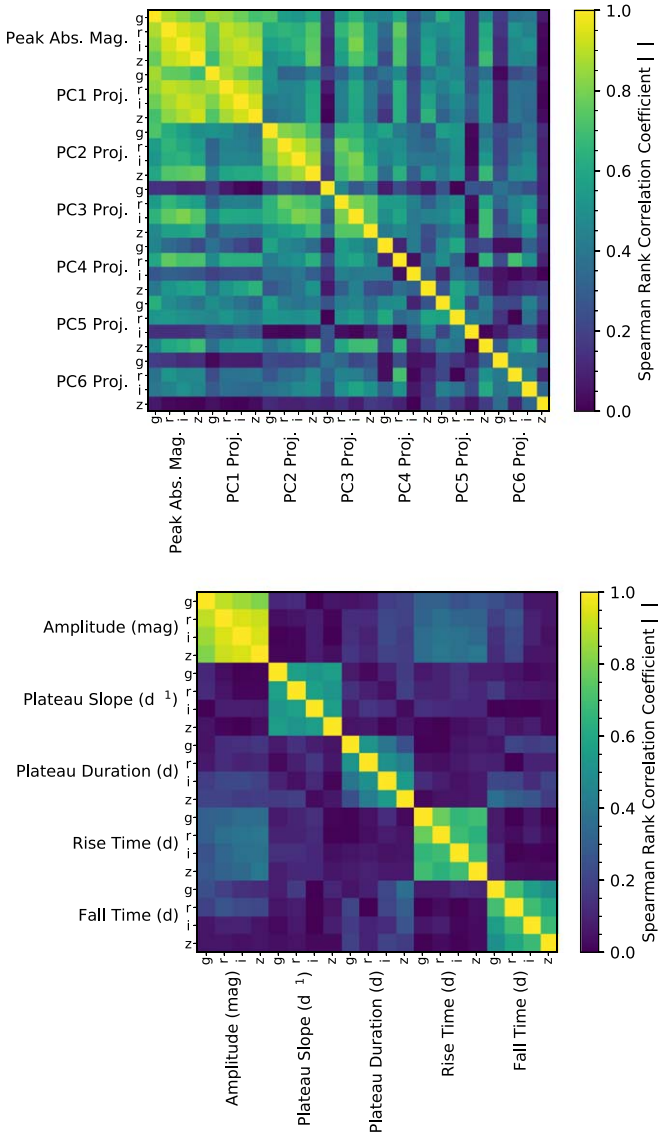

Figure 11. Top: absolute values of the Spearman rank correlation coefficients between each pair of features in our training set. Note the strong correlations between the four filters, as well as a correlation between the peak absolute magnitude and the projection onto the first principal component of the light curves. Correlations among the remaining features are weak. Bottom: the equivalent correlation matrix for the model parameters.

which is consistent with zero for all filters. The results are shown in Figure 12.

In all cases, the peak absolute magnitude and the projection onto the first principal component—we cannot compare these to each other because they are correlated—are by far the most discriminating between the classes. The remaining principal component coefficients contribute roughly in order of their rank. We again emphasize that cross-filter comparisons are meaningless in our analysis, but the relative importances of the features are similar for each filter.

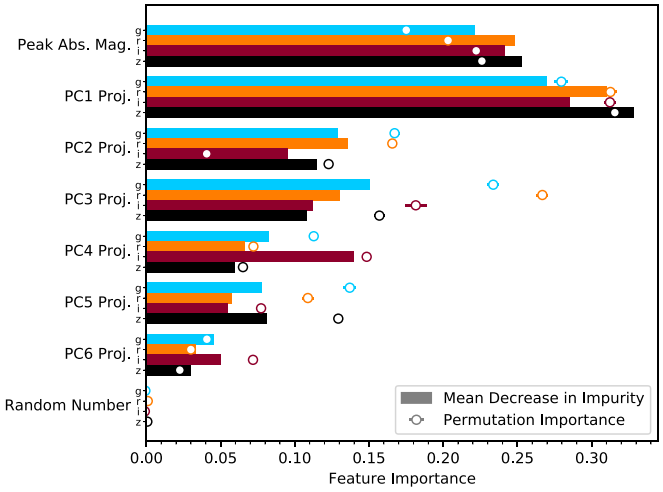

Figure 12. Feature importances for our training set. Cross-filter comparisons, as well as comparison between the peak absolute magnitude and the projection onto the first principal component, are meaningless due to the correlations at left. Nonetheless, it is clear that peak absolute magnitude and the projection onto the first principal component are the most important in all filters.

Appendix D
Hyperparameter Optimization

Our random forest classifier has several hyperparameters that can be adjusted to obtain better results. We repeated our analysis, apart from the final classification, over a grid of four of these parameters (represented by their variable names in scikit-learn; Pedregosa et al. 2011):

1. `criterion` ∈ {Gini impurity, entropy}, the function to measure the quality of a split;
2. `max_depth` ∈ {5, 10, 15, 20, 25, 29}, the maximum depth of a decision tree;
3. `max_features` ∈ {5, 10, 15, 20, 25}, the number of features to consider when looking for the best split; and
4. `n_estimators` ∈ {10, 20, 50, 100, 200, 500}, the number of decision trees in the random forest.

For each of these 360 classifiers, we calculated 12 different metrics on our training set: completeness and purity in each of our 5 classes, accuracy, and F_1 score (see Appendix A for definitions). Figure 13 shows the results when using entropy as the split criterion; switching to the Gini impurity was neutral or slightly worse in most cases. There is no clear winner in all 12 metrics, but in general, we find it important not to limit the depth of the decision trees. The final classifier used in our analysis uses an entropy criterion, no maximum depth, a maximum of 5 features, and 100 estimators.

In addition to optimizing the random forest (RF) hyperparameters, we repeated our analysis using two other classification algorithms: a support vector machine (SVM; with regularization parameter $C = 1000$ and kernel coefficient $\gamma = 0.1$) and a multilayer perceptron (MLP; with two hidden layers of 10 and 5 neurons and an L2 penalty parameter $\alpha = 10^{-5}$). We also repeated our analysis using the model parameters directly as features, rather than performing a PCA. Table 6 lists the same 12 metrics for our baseline procedure (which has the highest F_1 score) in the top row and how these metrics change using the other classifiers and feature set. The RF with the model parameters and the SVM perform similarly to our baseline algorithm, but with some important weaknesses (e.g., low SLSN completeness and purity, respectively). The MLP performs worse in most metrics.

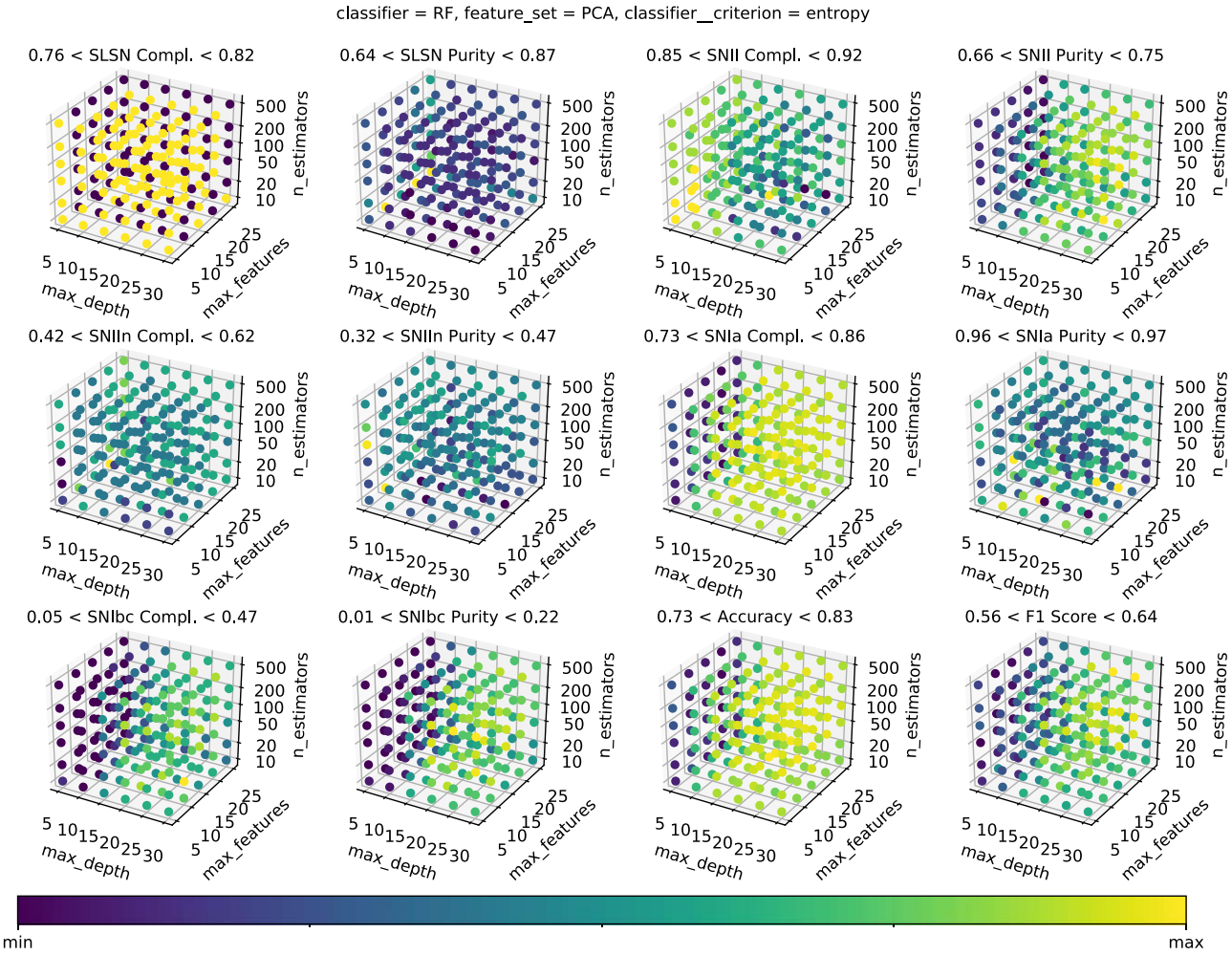

Figure 13. Results of varying the hyperparameters of our classifier over a three-dimensional grid, as measured by 12 metrics. The range of each metric is shown above each plot, and the color of each point corresponds to where it lies in that range. There is no single set of hyperparameters that optimizes all of the metrics, but in general, we find it important not to limit the maximum depth of the decision trees.

(The data used to create this figure are available.)

Table 6
Other Classifiers and Feature Sets

Classifier	Feature Set	Completeness					Purity					Accuracy	F_1 Score
		SLSN	SN II	SN IIn	SN Ia	SN Ibc	SLSN	SN II	SN IIn	SN Ia	SN Ibc		
RF	PCA	0.824	0.892	0.500	0.849	0.368	0.667	0.728	0.387	0.961	0.206	0.824	0.628
RF	parameters	−0.176	−0.075	+0.125	−0.010	+0.105	+0.067	+0.024	−0.133	+0.002	+0.094	−0.016	−0.009
SVM	PCA	+0.118	+0.011	−0.125	+0.012	−0.053	−0.326	+0.050	+0.022	−0.002	+0.147	+0.007	−0.035
MLP	PCA	−0.176	−0.043	−0.083	−0.045	+0.053	−0.020	−0.016	−0.143	+0.012	−0.058	−0.047	−0.062

ORCID iDs

Griffin Hosseinzadeh https://orcid.org/0000-0002-0832-2974
Frederick Dauphin https://orcid.org/0000-0001-5123-6388
V. Ashley Villar https://orcid.org/0000-0002-5814-4061
Edo Berger https://orcid.org/0000-0002-9392-9681
David O. Jones https://orcid.org/0000-0002-6230-0151
Ryan Chornock https://orcid.org/0000-0002-7706-5668
Maria R. Drout https://orcid.org/0000-0001-7081-0082
Robert P. Kirshner https://orcid.org/0000-0002-1966-3942
Ragnhild Lunnan https://orcid.org/0000-0001-9454-4639
Raffaella Margutti https://orcid.org/0000-0003-4768-7586
Dan Milisavljevic https://orcid.org/0000-0002-0763-3885
Yen-Chen Pan https://orcid.org/0000-0001-8415-6720
Armin Rest https://orcid.org/0000-0002-4410-5387
Daniel M. Scolnic https://orcid.org/0000-0002-4934-5849
Eugene Magnier https://orcid.org/0000-0002-7965-2815
Nigel Metcalfe https://orcid.org/0000-0001-9034-4402
Richard Wainscoat https://orcid.org/0000-0002-1341-0952
Christopher Waters https://orcid.org/0000-0003-1989-4879

References

Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Algeo, J. 1977, American Speech, 52, 47

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Baldeschi, A., Miller, A., Stroh, M., Margutti, R., & Coppejans, D. L. 2020, ApJ, 902, 60
Balestra, I., Mainieri, V., Popesso, P., et al. 2010, A&A, 512, A12
Barbary, K. 2016, extinction v0.3.0, Zenodo, doi:10.5281/zenodo.804967
Bayes, M., & Price, M. 1763, RSPT, 53, 370
Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002
Bianco, F. B., Modjaz, M., Hicken, M., et al. 2014, ApJS, 213, 19
Boone, K. 2019, AJ, 158, 257
Breiman, L. 2001, MachL, 45, 5
Bronder, T. J., Hook, I. M., Astier, P., et al. 2008, A&A, 477, 717
Cannon, R., Drinkwater, M., Edge, A., et al. 2006, MNRAS, 372, 425
Cappellaro, E., Benetti, S., Pastorello, A., et al. 2012, CBET, 3274, 1
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
Charnock, T., & Moss, A. 2017, ApJL, 837, L28
Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2002, Journal of Artificial Intelligence Research, 16, 321
Chornock, R., Berger, E., Gezari, S., et al. 2014, ApJ, 780, 44
Colless, M., Peterson, B. A., Jackson, C., et al. 2003, arXiv:astro-ph/0306581
Cowie, L. L., Barger, A. J., & Hu, E. M. 2010, ApJ, 711, 928
da Costa-Luis, C. 2019, JOSS, 4, 1277
Dressler, A., & Gunn, J. E. 1992, ApJS, 78, 1
Drinkwater, M. J., Jurek, R. J., Blake, C., et al. 2010, MNRAS, 401, 1429
Drout, M. R., Chornock, R., Soderberg, A. M., et al. 2014, ApJ, 794, 23
Elias, J. H., Matthews, K., Neugebauer, G., & Persson, S. E. 1985, ApJ, 296, 379
Finkelstein, S. L., Cohen, S. H., Malhotra, S., et al. 2009, ApJL, 703, L162
Fitzpatrick, E. L. 1999, PASP, 111, 63
Foley, R. J., & Mandel, K. 2013, ApJ, 778, 167
Fremling, U. C., Miller, A. A., Sharma, Y., et al. 2019, arXiv:1910.12973
Gal-Yam, A. 2012, Sci, 337, 927
Gal-Yam, A. 2016, in Handbook of Supernovae, ed. A. W. Alsabti & P. Murdin (Cham: Springer)
Gal-Yam, A. 2019, ARA&A, 57, 305
Garcet, O., Gandhi, P., Gosset, E., et al. 2007, A&A, 474, 473
Gelman, A., & Rubin, D. B. 1992, StaSc, 7, 457
Gezari, S., Chornock, R., Rest, A., et al. 2012, Natur, 485, 217
Gomez, S., Berger, E., Blanchard, P. K., et al. 2020, ApJ, 904, 74
Graham, M. L., Connolly, A. J., Ivezić, Ž, et al. 2018, AJ, 155, 1
Graur, O., Bianco, F. B., Huang, S., et al. 2017a, ApJ, 837, 120
Graur, O., Bianco, F. B., Modjaz, M., et al. 2017b, ApJ, 837, 121
Hasinger, G., Capak, P., Salvato, M., et al. 2018, ApJ, 858, 77
Hastings, W. K. 1970, Biometrika, 57, 97
Hewett, P. C., & Wild, V. 2010, MNRAS, 405, 2302
Holoien, T. W. S., Brown, J. S., Vallely, P. J., et al. 2019, MNRAS, 484, 1899
Hosseinzadeh, G., & Dauphin, F. 2020, Superphot, Zenodo, doi:10.5281/zenodo.374789
Hosseinzadeh, G., Dauphin, F., Villar, V. A., et al. 2020, Host Galaxy Spectra of Pan-STARRS1 Transients, Zenodo, doi:10.5281/zenodo.3973698
Howell, D. A., Sullivan, M., Nugent, P. E., et al. 2006, Natur, 443, 308
Hunter, J. D. 2007, CSE, 9, 90
Im, M., Faber, S. M., Gebhardt, K., et al. 2001, AJ, 122, 750
Ishida, E. E. O., Beck, R., González-Gaitán, S., et al. 2019, MNRAS, 483, 2
Ivezić, Ž, Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jha, S., Riess, A. G., & Kirshner, R. P. 2007, ApJ, 659, 122
Jones, D. H., Read, M. A., Saunders, W., et al. 2009, MNRAS, 399, 683
Jones, D. O., Scolnic, D. M., Riess, A. G., et al. 2017, ApJ, 843, 6
Jones, D. O., Scolnic, D. M., Riess, A. G., et al. 2018, ApJ, 857, 51
Karhunen, K., Kotilainen, J. K., Falomo, R., & Bettoni, D. 2014, MNRAS, 441, 1802
Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009, PASP, 121, 1028
Kimura, A., Takahashi, I., Tanaka, M., et al. 2017, in 2017 IEEE 37th Int. Conf. Distributed Computing Systems Workshops (ICDCSW) (Piscataway, NJ: IEEE), 354
Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. A. 2019, JOSS, 4, 1143
Kurtz, M. J., & Mink, D. J. 1998, PASP, 110, 934
Lamareille, F., Brinchmann, J., Contini, T., et al. 2009, A&A, 495, 53
Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, A&A, 439, 845
Lemaître, G., Nogueira, F., & Aridas, C. K. 2017, Journal of Machine Learning Research, 18, 1
Lidman, C., Tucker, B. E., Davis, T. M., et al. 2020, MNRAS, 496, 19
Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, ApJS, 172, 70
Louppe, G. 2015, PhD thesis, Univ. Liège arXiv:1407.7502
Lunnan, R., Chornock, R., Berger, E., et al. 2014, ApJ, 787, 138
Lunnan, R., Chornock, R., Berger, E., et al. 2018, ApJ, 852, 81
Magnier, E. A., Chambers, K. C., Flewelling, H. A., et al. 2020a, ApJS, 251, 3
Magnier, E. A., Schlafly, E. F., Finkbeiner, D. P., et al. 2020b, ApJS, 251, 6
Magnier, E. A., Sweeney, W. E., Chambers, K. C., et al. 2020c, ApJS, 251, 5
Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, ApJ, 877, 81
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, JChPh, 21, 1087
Minkowski, R. 1941, PASP, 53, 224
Möller, A., Ruhlmann-Kleider, V., Leloup, C., et al. 2016, JCAP, 12, 008
Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, PASP, 131, 118002
Narayan, G., Foley, R. J., Berger, E., et al. 2011, ApJL, 731, L11
Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5
Norris, R. P., Afonso, J., Appleton, P. N., et al. 2006, AJ, 132, 2409
Oliphant, T. E. 2006, A Guide to NumPy (USA: Trelgol Publishing)
Owen, F. N., & Morrison, G. E. 2009, ApJS, 182, 625
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Perez, F., & Granger, B. E. 2007, CSE, 9, 21
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, 13
Quimby, R. M., Cia, A. D., Gal-Yam, A., et al. 2018, ApJ, 855, 2
Quimby, R. M., Kulkarni, S. R., Kasliwal, M. M., et al. 2011, Natur, 474, 487
Quimby, R. M., Werner, M. C., Oguri, M., et al. 2013, ApJL, 768, L20
Rest, A., Scolnic, D., Foley, R. J., et al. 2014, ApJ, 795, 44
Rest, A., Stubbs, C., Becker, A. C., et al. 2005, ApJ, 634, 1103
Richards, J. W., Homrighausen, D., Freeman, P. E., Schafer, C. M., & Poznanski, D. 2012, MNRAS, 419, 1121
Riess, A. G., Strolger, L.-G., Tonry, J., et al. 2004a, ApJL, 600, L163
Riess, A. G., Strolger, L.-G., Tonry, J., et al. 2004b, ApJ, 607, 665
Ross, N. P., Shanks, T., Cannon, R. D., et al. 2008, MNRAS, 387, 1323
Rovilos, E., Fotopoulou, S., Salvato, M., et al. 2011, A&A, 529, A135
Sako, M., Bassett, B., Connolly, B., et al. 2011, ApJ, 738, 162
Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PeerJ Computer Science, 2, e55
Sanders, N. E., Soderberg, A. M., Foley, R. J., et al. 2013, ApJ, 769, 39
Sanders, N. E., Soderberg, A. M., Gezari, S., et al. 2015, ApJ, 799, 208
Scarlata, C., Colbert, J., Teplitz, H. I., et al. 2009, ApJL, 704, L98
Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Schlegel, E. M. 1990, MNRAS, 244, 269
Smith, A. G., Hopkins, A. M., Hunstead, R. W., & Pimbblet, K. A. 2012, MNRAS, 422, 25
Spearman, C. 1904, The American Journal of Psychology, 15, 72
Sravan, N., Milisavljevic, D., Reynolds, J. M., Lentner, G., & Linvill, M. 2020, ApJ, 893, 127
Stalin, C. S., Petitjean, P., Srianand, R., et al. 2010, MNRAS, 401, 294
Stritzinger, M. D., Anderson, J. P., Contreras, C., et al. 2018, A&A, 609, A134
Szokoly, G. P., Bergeron, J., Hasinger, G., et al. 2004, ApJS, 155, 271
Taddia, F., Sollerman, J., Leloudas, G., et al. 2015, A&A, 574, A60
Taddia, F., Stritzinger, M. D., Bersten, M., et al. 2018, A&A, 609, A136
Tajer, M., Polletta, M., Chiappetti, L., et al. 2007, A&A, 467, 73
Theano Development Team 2016, arXiv:1605.02688
Tonry, J., & Davis, M. 1979, AJ, 84, 1511
Trump, J. R., Impey, C. D., Elvis, M., et al. 2009, ApJ, 696, 1195
Uomoto, A., & Kirshner, R. P. 1985, A&A, 149, L7
Valenti, S., Howell, D. A., Stritzinger, M. D., et al. 2016, MNRAS, 459, 3939
Villar, V. A., Berger, E., Miller, G., et al. 2019, ApJ, 994, 83
Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020a, ApJ, 905, 94
Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020b, Light Curves of Pan-STARRS1 SN-like Transients, Zenodo, doi:10.5281/zenodo.3974950
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Waters, C. Z., Magnier, E. A., Price, P. A., et al. 2020, ApJS, 251, 4
Wen, Z. L., & Han, J. L. 2015, ApJ, 807, 178
Wheeler, J. C., & Harkness, R. P. 1986, in Galaxy Distances and Deviations from Universal Expansion, ed. B. F. Madore & R. B. Tully (Dordrecht: Springer)
Wheeler, J. C., & Levreault, R. 1985, ApJL, 294, L17